STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 142280

To: Barba Koroma
Location: REM/2C18
Art Unit: 1638
Wednesday, January 26, 2005

Case Serial Number: 10/808979

From: Beverly Shears
Location: Remsen Bldg.
RM 1A54
Phone: 571-272-2528

beverly.shears@uspto.gov

Search Notes

This Page Blank (uspto)

```
                        GenCore version 5.1.6
               Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 24, 2005, 23:06:56 ; Search time 28265 Seconds
                                              (without alignments)
                                              11709.910 Million cell updates/sec

Title:          US-10-808-979-18
Perfect score:  6999
Sequence:       1 gcgatggcaatggaaaatgt..........gacgttttagacatgcaata 6999

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters:      9053458

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                SUMMARIES
                       %
Result                Query
   No.     Score      Match Length DB  ID                          Description
-----------------------------------------------------------------------------
     1    2732.8       39.0    6815  1  AB076662                    AB076662 Pantoea a
     2    2626.2       37.5    6918  1  ERWCRT                      D90087 Erwinia ure
     3    2624.6       37.5    6918  6  I12910                      I12910 Sequence 13
     4      2620       37.4    6586  1  AY166713                    AY166713 Pantoea s
     5    2468.8       35.3    6965  1  ERWCRTA                     M90698 Pantoea agg
     6    2135.2       30.5   12753  1  ERWCRTS                     M87280 Pantoea agg
     7    1900.6       27.2    4624  6  CQ793208                    CQ793208 Sequence
     8    1374.8       19.6  349107  1  BX571873                    BX571873 Photorhab
     9      1189       17.0    2415  1  ERWCRTIB                    M38423 Pantoea agg
    10    1179.8       16.9    3801  1  AF408848                    AF408848 Xanthobac
    11    1105.8       15.8    6635  1  AF218415                    AF218415 Bradyrhiz
    12    1041.2       14.9    6335  1  PMCRTWZYI                   Y15112 Paracoccus
    13    1037.2       14.8    5373  1  ATUCRTWA                    D58420 Paracoccus
c   14    1035.4       14.8    5188  1  FAU62808                    U62808 Flavobacter
c   15    1035.4       14.8    8625  6  AR169831                    AR169831 Sequence
c   16    1035.4       14.8    8625  6  AR452005                    AR452005 Sequence
c   17    1035.4       14.8   11233  6  AR169852                    AR169852 Sequence
c   18    1035.4       14.8   11233  6  AR452026                    AR452026 Sequence
    19     827.4       11.8    1493  6  BD241820                    BD241820 Method of
```

```
    20     827.4       11.8    1493  6  AX014697                    AX014697 Sequence
    21     826.2       11.8    1479  6  CQ793426                    CQ793426 Sequence
    22     826.2       11.8    1479  6  CQ793595                    CQ793595 Sequence
    23     826.2       11.8    1479  6  E22381                      E22381 beta-Carote
    24     826.2       11.8    1479  6  I12907                      I12907 Sequence 10
    25     808.6       11.6    1479  6  AX394988                    AX394988 Sequence
    26     713.4       10.2    1518  6  I72653                      I72653 Sequence 7
    27     708.4       10.1    1522  6  I72654                      I72654 Sequence 9
    28     619.4        8.8    1482  6  A58565                      A58565 Sequence 6
    29     619.4        8.8    1482  6  A84702                      A84702 Sequence 12
    30     619.4        8.8    1482  6  AR103081                    AR103081 Sequence
    31     619.4        8.8    1482  6  AR139697                    AR139697 Sequence
    32     619.4        8.8    1482  6  AR391876                    AR391876 Sequence
    33     604.4        8.6  349723  1  BX842650                    BX842650 Bdellovib
    34     515.6        7.4    1198  6  I24838                      I24838 Sequence 5
    35     515.6        7.4    1198  6  I72652                      I72652 Sequence 5
    36       479        6.8   30040  1  AF182374                    AF182374 Bradyrhiz
    37     467.4        6.7  349737  1  BX572597                    BX572597 Rhodopseu
    38     461.4        6.6    1581  1  D83514                      D83514 Erythrobact
    39     461.4        6.6    1581  6  E11103                      E11103 Gene coding
    40     449.6        6.4    1083  6  I40224                      I40224 Sequence 1
    41     449.6        6.4    5397 12  AY605097                     AY605097 Synthetic
    42     449.2        6.4    1296  6  AX394984                    AX394984 Sequence
    43       438        6.3    1235  6  I72656                      I72656 Sequence 13
    44       430        6.1    1232  6  AR222522                    AR222522 Sequence
    45       430        6.1    1232  6  AR432935                    AR432935 Sequence

                               ALIGNMENTS

RESULT 1
AB076662
LOCUS       AB076662                 6815 bp    DNA     linear   BCT 26-DEC-2001
DEFINITION  Pantoea agglomerans pv. milletiae crtE, crtX, crtY, crtI, crtB,
            crtZ genes, complete cds.
ACCESSION   AB076662
VERSION     AB076662.1  GI:18143444
KEYWORDS    .
SOURCE      Pantoea agglomerans pv. milletiae
  ORGANISM  Pantoea agglomerans pv. milletiae
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1
  AUTHORS   Kamiunten,H. and Hirata,R.
  TITLE     Isolation and characterization of carotenoid biosynthesis genes
            from Pantoea agglomerans pv. milletiae Wist 801
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6815)
  AUTHORS   Kamiunten,H. and Hirata,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-DEC-2001) Hiroshi Kamiunten, Miyazaki University,
            Faculty of Agriculture; Gakuenkibanadai Nishi 1-1, Miyazaki,
            Miyazaki 889-2155, Japan (E-mail:a01108u@cc.miyazaki-u.ac.jp,
            Tel:81-985-58-7170, Fax:81-985-58-7170)
FEATURES             Location/Qualifiers
     source          1..6815
                     /organism="Pantoea agglomerans pv. milletiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:182454"
                     /note="pathovar: milletiae"
     gene            160..1071
                     /gene="crtE"
     CDS             160..1071
                     /gene="crtE"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB79600.1"
                     /db_xref="GI:18143445"
                     /translation="MMTVCAEQHVNFIHSDAASLLNDIEQRLDQLLPVESERDLVGAA
                     MRDGALAPGKRIRPLLLLLAARDLGCNATPAGLLDLACAVEMVHAASLILDDMPCMDD
                     AQLRRGRPTIHCQYGEHVAILAAVALLSKAFGVVAAAEGLTATARADAVAELSHAVGM
                     QGLVQGQFKDLSEGDKPRSADAILMTNHYKTSTLFCASMQMASIVAEASGEAREQLHR
                     FSLNLGQAFQLLDDLTDGMADTGKDAHQDDGKSTLVNLLGPQAVETRLRDHLRCASEH
                     LLSACQDGYATHHFVQAWFEKKLAAVS"
     gene            1081. .2376
                     /gene="crtX"
     CDS             1081. .2376
                     /gene="crtX"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB79601.1"
                     /db_xref="GI:18143446"
                     /translation="MSHFAVIAPPFYSHVQALQHLSQALIARGHQITFIQQTDVSALL
                     TDSRIGFFPLGLASHPAGSLAHTLQLAAHPLGPSMLKLINEMARSTDMLCRELPVVLS
                     TLAIDGVIVDQMEPAGALAAEALNLPYVSVACALPLNREADFPLAVMPFDYARTDQAR
                     ERYRTSEKIYDWLMRRHDRVIARNAHAMGLAPREKLHHCFSPVAQINQLIPELDFPRQ
                     TLPDHFHSVGPLRTTEPAPAASQPRYFPHGDRPRIFASLGTLQGHRYGLFKTIVRACQ
                     EIDAQLLLAHCGRLSPFQAEKLAQASHVQVVDFADQAAALAQADLVITHGGMNTVLDG
                     INHLTPLLTIPLAFDQPGVAARVVWHGIGRRASRFTTSHSMARQLQTLLADESYVQQM
                     KTLRSALRQAGGTTLAADIVEQAMLTRQPVLTRRHYAAV"
     gene            2363. .3523
                     /gene="crtY"
     CDS             2363. .3523
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB79602.1"
                     /db_xref="GI:18143447"
                     /translation="MPRYDLILVGAGLANGLIALRLRQQRPSLRILLIDAEREPGANH
                     TWSFHAEDLTETQHRWIAPLVVHHWPGYEVRFPQRSRSMNSGYFCVTAERFAQVIRDR
                     FAPDLLLNTRVASIASRAVTLDDGRVLESDAVIDGRGYQPDGALRMGFQSFVGQEWQL
                     SEPHGLTAPIIMDATVDQQAGYRFVYSLPFSADTLLIEDTHYIDNATLEGDRARQNIH
                     DYAAQQGWRLDRLLREEQGALPITLTGDVAAFWQNRDLPCSGLRAGLFHPTTGYSLPL
                     AVALADRLAQMQTFTSETLHATIQQFASQAWQQQRFFRMLNRMLFLAGPADQRWQVMQ
                     RFYGLPEGLIARFYAGKLTLPDRLRILSGKPPVPVLAALQAIMTPHRQQAMQ"
     gene            3520. .4998
                     /gene="crtI"
     CDS             3520. .4998
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB79603.1"
                     /db_xref="GI:18143448"
                     /translation="MNRTTVIGAGFGGLALAIRLQASGVPTRLLEQRDKPGGRAYVYQ
                     DQGFTFDAGPTVITDPSAIEELFTLAGKKLSDYVELMPVKPFYRLCWESGKVFSYDND
                     QPALEAQIAAFNPRDVEGYRRFLAYSRAVFAEGYLKLGTVPFLSFRDMLRAAPQLAKL
                     QAWRSVYSKVASYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLIHALEREWGVWFPRG
                     GTGALVQGMVKLFEDLGGEVELNASVARLETQENRITAVHLKDGRVFPTRAVASNADV
                     VHTYRELLSQHPASQAQGRSLQNKRMSNSLFVIYFGLNHHHNQLAHHTVCFGPRYREL
                     IDEIFNKDGLAEDFSLYLHAPCVTDPSLAPEGCGSYYVLAPVPHLGTADIDWAVEGPR
                     LRDRIFDYLEQHYMPGLRSQLVTHRIFTPFDFRDELNAYQGSAFSVEPILTQSAWFRP
                     HNRDKNINNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLA"
     gene            5034. .5924
                     /gene="crtB"
     CDS             5034. .5924
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB79604.1"
                     /db_xref="GI:18143449"
                     /translation="MEVGSKSFATASKLFGAKTRRSVLMLYAWCRHCDDVIDDQVLGF
                     SNDTPSLQSAEQRLAQLEMKTRQAYAGSQMHEPAFAAFQEVAMAHDILPAYAFDHLAG
                     FAMDVHETRYQTLDDTLRYCYHVAGVVGLMMAQIMGVRDNATLDRACDLGLAFQLTNI
                     ARDIVEDADAGRCYLPATWLAEEGLTRENLADPQNRQALSRVARRLVETAEPYYRSAS
                     AGLPGLPLRSAWAIATAQQVYRKIGMKVVQAASQAWDQRQSTSTPEKLALLVAASGQA
                     VTSRVARHAPRSADLWQRPV"
     gene            complement(5863. .6390)
                     /gene="crtZ"
     CDS             complement(5863. .6390)
                     /gene="crtZ"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAB79605:1"
                     /db_xref="GI:18143450"
                     /translation="MLWIWNALIVLVTVIGMEITAALAHRYIMHGWGWGWHLSHHEPH
```

```
                     KGWFEVNDLYAVVFAALSILLIYLGSTGVWPLQWIGAGMTLYGLLYFIVHDGLVHQRW
                     PFRYVPRRGYLRRLYMAHRMHHAVRGKEGCVSFGFLYAPPLSKLQATLRERHGVKRGA
                     ARDQRSVERDAPPGK"
ORIGIN

  Query Match             39.0%;  Score 2732.8;  DB 1;  Length 6815;
  Best Local Similarity   65.8%;  Pred. No. 0;
  Matches 4085;  Conservative    0;  Mismatches 2092;  Indels   27;  Gaps    7;

Qy        376 CTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGCGAT 435
              |||  | |     |  ||   |  ||||| ||  |  | || |  | |    || | ||
Db        217 CTGTTGAACGACATTGAGCAACGGCTTGATCAGCTTTTACCGGTTGAAAGCGAACGTGAC 276

Qy        436 CGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCTTTA 495
                 ||| | || ||  || |||||  | |||  ||||||| |  || || ||||| |||||  | 
Db        277 TTAGTGGGCGCTGCCATGCGCGACGGTGCGCTGGCACCAGGAAAGCGTATCCGTCCACTG 336

Qy        496 TTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTCGAT 555
               |  || |||||||||||||||||| |||| ||| |    ||||     ||| | ||||||
Db        337 CTGTTGTTGCTGGCAGCGCGCGATCTGGGCTGCAACGCCACGCCTGCCGGCCTGCTCGAT 396

Qy        556 CTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCG 615
              |||||||| || || || |||||||| || ||||| ||||||||||||||||| ||||  
Db        397 CTCGCCTGCGCGGTAGAGATGGTGCATGCCGCATCACTGATTCTGGATGACATGCCCTGC 456

Qy        616 ATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAAC 675
              |||||| | |||||  ||||||| || || || ||| | ||| || | | ||||||| | 
Db        457 ATGGATGATGCGCAACTGCGTCGCGGACGTCCGACCATTCATTGCCAGTATGGTGAACAT 516

Qy        676 GTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCA 735
              || |||||||| |||||  | || ||||| ||    ||||| |  ||| | ||    |||
Db        517 GTCGCGATTCTGGCCGCGGTGGCCCTGCTGAGTAAGGCATTCGGCGTGGTCGCTGCGGCA 576

Qy        736 CCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGC 795
                 || ||  | || |     | | | || ||  | || ||| | |||   || ||||||
Db        577 GAAGGCTTAACGGCAACCGCCAGAGCCGACGCTGTGGCAGAATTATCCCACGCAGTCGGC 636

Qy        796 CTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGC 855
               |||||||  | ||||| ||||||||  ||||||||   ||| ||     ||   ||||||
Db        637 ATGCAGGGGCTGGTGCAGGGGCAATTTAAGGATCTCTCCGAGGGTGACAAGCCACGCAGC 696

Qy        856 CCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTG 915
               | || || ||    ||||||||  |    |||||||||  ||||||  ||||| |  ||
Db        697 GCTGACGCCATTCTGATGACCAATCACTATAAAACCAGCACGCTGTTCTGCGCCTCCATG 756

Qy        916 CAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTC 975
              || |||||  | || |  ||||| || |||    | |  ||  ||   ||   |   || 
Db        757 CAGATGGCCTCTATCGTGGCTGAAGCCTCAGGTGAAGCCCGCGAACAGCTGCACCGTTTT 816

Qy        976 GCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAACAC 1035
               | |   || | || ||||| || || || || ||||| ||| | |||||        ||
Db        817 TCGCTTAATCTTGGTCAGGCTTTCCAGCTACTGGACGATCTCACTGACGGCATGGCCGAC 876

Qy       1036 ACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCT 1095
              |||||||||||||  || ||||| | || ||||| ||||||||  |  |||| ||  | 
Db        877 ACCGGTAAAGATGCCCATCAGGATGACGGGAAATCAACGCTGGTGAATCTGCTGGGACCA 936

Qy       1096 GACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCCTGC 1155
               | |  |  |||   || ||||||||||| |||||| ||||    |  || ||    |  
Db        937 CAGGCGGTTGAAACGCGACTGCGCGATCATCTGCGCTGCGCCAGCGAGCATCTGTTATCG 996

Qy       1156 GCCTGCCATCGCGGCATCGCCACTCGCCAATATATGCACGCGCTGTTTAATCAACAGCTA 1215
              ||||||||   |||    ||||| | ||| | | | || ||   |||| |  || | || 
Db        997 GCCTGCCAGGACGGTTATGCCACACACCATTTTGTTCAGGCCTGGTTTGAGAAAAAACTC 1056

Qy       1216 GCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGCCGCCGC 1275
              ||     |||  | ||   |   ||     || ||||||||||||| || || |||||  
Db       1057 GCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCACTTTGCGGTCATTGCACCGCCCT 1111

Qy       1276 TCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCGGCCATC 1335
              ||||||||||  | || ||| | |||    |    ||  |  |  | || ||||| ||||
Db       1112 TCTACAGCCATGTGCAGGCGCTTCAGCACCTTAGCCAGGCCTTAATCGCACGCGGACATC 1171
```

```
Qy          1336 GCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCATCGATT 1395
                   || || |||||||||||  | || |   || |  | || | |||    || ||||  |
Db          1172 AGATTACTTTCATCCAGCAGACGGACGTTAGCGCGCTACTCACCGATAGCCGTATCGGCT 1231

Qy          1396 TTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGCATCGGC 1455
                 || |  |  | ||   |    ||||||| || ||    ||||| | |    |||| | |
Db          1232 TTTTCCCGCTGGGTTTAGCCTCGCATCCGGCGGGCAGTCTGGCACACACCCTGCAACTGG 1291

Qy          1456 TGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCG 1515
                  |||    ||   |||||  ||  || |     ||||| | ||  | ||   | ||||||
Db          1292 CGGCGCATCCTCTCGGCCCCTCAATGCTGAAACTGATCAATGAGATGGCCCGCAGCACCG 1351

Qy          1516 ATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCGTGATCG 1575
                 ||||||| || || |||||||| | ||| ||||   |  ||   || ||||||||||| |
Db          1352 ATATGCTCTGTCGTGAACTGCCCGTGGTGCTGAGCACCCTGGCGATTGATGGCGTGATTG 1411

Qy          1576 CCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTT 1635
                  |||  | |||||  | || || | | |||  || ||||||||| | ||||| | |||||
Db          1412 TCGATCAGATGGAGCCTGCCGGTGCACTGGCGGCAGAAGCGCTGAACCTGCCTTATGTTT 1471

Qy          1636 CGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCT 1695
                 |||| ||||| ||  ||||  | || |||||||| |   | ||||| || |||||||| |
Db          1472 CGGTTGCCTGTGCGCTGCCACTTAACCGTGAAGCGGATTTCCCGCTGGCCGTGATGCCGT 1531

Qy          1696 TCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATG 1755
                 |    | |||   |   ||  | || | | |||| | ||   ||||     | |||||||
Db          1532 TTGACTATGCCAGGACCGATCAGGCACGGGAACGCTATCGCACCAGTGAAAAAATCTATG 1591

Qy          1756 ATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGA 1815
                 |  |  | ||||| |||||||     ||||||  |    |||| |  ||| |    ||
Db          1592 ACTGGCTGATGCGACGTCACGATCGGGTGATCGCCCGCAACGCTCATGCGATGGGATTAG 1651

Qy          1816 CGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGC 1875
                 |    |||     | | ||||| ||  | || ||| |||| || |||| |||| || | |
Db          1652 CACCACGGGAAAAACTGCATCACTGTTTTTCACCGGTGGCGCAGATCAACCAGCTGATTC 1711

Qy          1876 CGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCC 1935
                 |||   | |||||||| || |||  ||||||||     |  ||  | || || || || |
Db          1712 CGGAACTGGATTTTCCCCGCCAGACACTGCCCGATCATTTCCATTCGGTTGGCCCGCTGC 1771

Qy          1936 GCGCC------CCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTCAGC 1989
                 |  ||      ||||   |  | ||  | |  | |     ||  ||             |
Db          1772 GTACCACTGAACCGGCACCCGCCGCATCTCAGCCTCGCTACTTCCCGCATGGCGACAGAC 1831

Qy          1990 CGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGCATC 2049
                 ||    ||| ||| || || || ||||||||| || |||||||| | | ||||||
Db          1832 CGCGCATTTTTGCTTCACTCGGCACGCTGCAGGGACATCGCTACGGCCTGTTTAAAACCA 1891

Qy          2050 TGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGATTAA 2109
                 | |  |  || ||||   || |         | |  || | || ||||| ||  | ||
Db          1892 TCGTCCGCGCCTGCCAGGAGATCGATGCTCAGTTATTGCTGGCGCATTGCGGTCGGTTGT 1951

Qy          2110 ACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCGTCG 2169
                    |    ||| || |  | |||| ||  ||  ||   |    ||||   ||||| | ||
Db          1952 CACCTTTCCAGGCGGAGAAACTGGCGCAGGCCAGCCATGTTCAGGTGGTCGATTTTGCCG 2011

Qy          2170 ATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACAGCG 2229
                 | |||   || || ||   ||| ||  | ||  | || || || | |||  | || |  |
Db          2012 ACCAGGCGGCTGCACTGGCGCAGGCCGATCTTGTCATTACGCACGGCGGCATGAATACGG 2071

Qy          2230 CGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATCAGC 2289
                  |||||| |   |  |   |   |||||| ||||| || | ||| | || ||||||||||
Db          2072 TGCTGGATGGGATTAACCACCTGACGCCGCTGCTGACGATCCCGCTGGCCTTTGATCAGC 2131

Qy          2290 CCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTAGCC 2349
                 | || ||||| ||  |  | |  ||||| |   | || || ||||| |||||||||| |
Db          2132 CGGGTGTGGCTGCCAGAGTGGTCTGGCACGGTATCGGACGTCGCGCCTCACGCTTTACCA 2191

Qy          2350 GTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGCTAC 2409
                      ||     |||  |  || || ||     | ||| | || ||  | ||||  |  |
Db          2192 CCAGCCACTCCATGGCACGTCAGCTCCAGACCTTACTGGCTGATGAAAGCTACGTTCAGC 2251

Qy          2410 GGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCGACA 2469
                  ||||  | || | |        || |  |  || |||||   |   |  || || ||||
Db          2252 AGATGAAAACGCTGCGTAGCGCACTTCGTCAGGCGGGCGGAACCACCCTGGCGGCGGACA 2311

Qy          2470 TCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCAC 2529
                 ||||||| |||||| ||      |  ||    |||||  |  ||| |||   ||||   |
Db          2312 TCGTCGAACAGGCGATGCTGACCCGCCAGCCGGTGCTCAC-CAGGAGACACTATGC---C 2367

Qy          2530 GCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCT 2589
                 ||  || ||| ||||| |||| ||||| |||||||| || ||  |||| || ||||| ||
Db          2368 GCGGTATGATCTGATTCTGGTAGGTGCGGGACTGGCCAACGGGCTGATCGCCCTGCGCCT 2427

Qy          2590 GCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGG 2649
                 | | ||   || |||    |||      ||| || | ||   |||     | ||||  |
Db          2428 GAGACAGCAGCGGCCCTCCCTGCGCATTCTGCTGATTGACGCCGAACGTGAACCGGGTGC 2487

Qy          2650 CAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCA 2709
                 ||||||||||||||||||||||      ||||||| ||     || | ||||||| |
Db          2488 CAATCATACCTGGTCGTTTCATGCGGAAGATCTCACCGAAACGCAGCATCGCTGGATCGC 2547

Qy          2710 ACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAA 2769
                  |||||| |       |  ||| | || ||| |||| || |||||    |   | ||||
Db          2548 TCCGCTGGTAGTACATCACTGGCCTGGCTATGAGGTCCGCTTTCCCCAACGCAGTCGCAG 2607

Qy          2770 TCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGC 2829
                 | || || | |  |||| || | |  |  | |     || || |      |   ||
Db          2608 TATGAACAGTGGCTATTTTTGCGTGACTGCGGAGCGCTTCGCGCAGGTCATCCGCGACAG 2667

Qy          2830 GATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGAC 2889
                 | | |     ||||||  |   || ||    ||        |     |      ||| ||
Db          2668 GTTTGCGCCGGATCTGCTGCTGAATACCCGGGTTGCGAGCATCGCTTCACGCGCGGTCAC 2727

Qy          2890 GCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCC 2949
                 |||||  |||||||| |  || |      | |  ||||||||||| || |||   |||||
Db          2728 GCTGGACGATGGCCGGGTGCTGGAGAGTGACGCAGTGATTGATGGCCGTGGCTACCAGCC 2787

Qy          2950 GACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCA 3009
                           ||||   |||| |  |||  |||| | ||||| |||||||||||     |
Db          2788 CGATGGCGCGCTGCGCATGGGCTTCCAGTCGTTTGTCGGACAGGAGTGGCAGCTCAGCGA 2847

Qy          3010 GCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTA 3069
                 ||| || ||||||     ||||||| | ||||||||||| || |||||||| || || ||
Db          2848 GCCACATGGCCTGACCGCGCCGATCATCATGGATGCCACGGTAGATCAGCAGGCAGGCTA 2907

Qy          3070 TCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTA 3129
                 ||| ||||||||||  |||||  |    || ||    ||  |||||||||||||||| ||
Db          2908 TCGCTTTGTCTACAGCCTGCCCTTTTCTGCAGACACCCTGCTGATTGAAGATACCCACTA 2967

Qy          3130 CGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAA 3189
                 | || |  |  || ||||||  |   |   ||| |||||| | |||  ||||||||||
Db          2968 CATTGATAACGCCACGCTGGAAGGCGATCGCGCCCGTCAGAATATCCACGACTATGCCGC 3027

Qy          3190 TCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTAC 3249
                  ||||| || |||    |      |||||| || ||||||||| |||   ||||| |||||
Db          3028 GCAGCAGGGATGGCGATTAGACCGGCTGCTACGCGAAGAGCAAGGCGCGTTACCCATTAC 3087

Qy          3250 CCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCT 3309
                  |||| |||| || | |     ||||||   ||| | ||||      | ||||||||  |
Db          3088 GCTGACCGGCGACGTGGCCGCCTTCTGG---CAGAACCGCGATTTGCCCTGCAGCGGTTT 3144

Qy          3310 GCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGC 3369
                 ||| |||||| | || ||| |||| ||||| || ||  ||||    || || || || ||
Db          3145 GCGTGCCGGGTTATTCCATCCCACTACCGGCTATTCTCTGCCACTGGCGGTTGCTCTGGC 3204

Qy          3370 GGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACG 3429
                  ||   | | ||   | ||  | |        |    || ||   |      ||  | |
Db          3205 TGACCGGCTGGCGCAGATGCAGACGTTTACCTCTGAGACCCTGCACGCCACCATTCAGCA 3264

Qy          3430 CTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTT 3489
                  || ||| |||||   ||||   ||||||| || |||||  |||| ||||| ||||||||
Db          3265 ATTCGCCAGTCAGGCCTGGCAGCAACAGCGCTTCTTCCGCATGCTGAACCGTATGCTGTT 3324

Qy          3490 TTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGC 3549
                 | |||| |||  | |  | ||||||||||  || ||||||||||||||| | ||    |
Db   3325 TCTGGCGGGTCCGGCTGACCAGCGCTGGCAGGTTATGCAACGTTTTACGGCCTTCCCGA 3384

Qy   3550 CGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCT 3609
           ||  | ||   ||| |||||| || ||  ||||     |||  |||     ||| ||  |
Db   3385 AGGTCTGATCGCCCGTTTTTATGCGGGAAAACTCACTTTGCCTGATCGGCTGCGCATCTT 3444

Qy   3610 GTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGA 3669
            |||||||||| ||||| ||| |    |  ||| | |  ||  |  ||| | ||  |
Db   3445 AAGCGGCAAGCCCCCGGTCCCCGTTCTGGCGGCGTTACAGGCTATTATGACTCCTCACCG 3504

Qy   3670 ACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCG 3729
           | |  |  ||   ||||||  | |||   || |||||||||||||||||| || |||||
Db   3505 TCAACAGGCGATGCAATGAATAGAACTACAGTAATTGGCGCAGGCTTTGGTGGTCTGGCT 3564

Qy   3730 CTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAA 3789
          ||||| |||||||| || ||| | ||| | || ||     | || |||||||| |||||
Db   3565 CTGGCAATTCGCCTTCAGGCGTCAGGCGTTCCCACCCGACTGCTGGAGCAGCGTGACAAG 3624

Qy   3790 CCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACG 3849
          || ||||| || || ||||| | | ||||    ||||| || || ||||||||| ||||||
Db   3625 CCTGGCGGCCGGGCTTATGTCTATCAGGATCAGGGCTTCACGTTTGATGCCGGCCCCACG 3684

Qy   3850 GTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGC 3909
          || ||||||||||||||||||||| |||||| ||||||| || || || ||  ||||
Db   3685 GTAATCACCGATCCCAGCGCCATTGAAGAGCTGTTCACCCTCGCGGGTAAAAAGCTCTCT 3744

Qy   3910 GATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAA 3969
          || || ||||||||||||||||| | ||| || ||||||||| |||||||||    |||||
Db   3745 GACTATGTCGAGCTGATGCCGGTGAAGCCGTTTTATCGCCTCTGCTGGGAGTCCGGCAAG 3804

Qy   3970 CAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCG 4029
            | |   ||| |||||  |||||||   ||||||   ||||||| ||| | ||| ||||||
Db   3805 GTGTTCAGTTATGACAACGATCAGCCCGCGCTGGAAGCGCAGATTGCCGCGTTTAATCCG 3864

Qy   4030 CAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGT 4089
          |  || || ||||| |||||||  ||||| |||||||| || |  || |||   || || 
Db   3865 CGTGACGTTGAAGGATATCGTCGCTTTCTGGCCTATTCCCGAGCGGTGTTTGCTGAAGGC 3924

Qy   4090 TATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCG 4149
          |||||||| || ||||| ||||||||||||    | || ||||||||||| | |||||||
Db   3925 TATCTGAAGCTTGGCACCGTGCCGTTTCTGTCATTCCGCGACATGCTGCGGGCCGCGCCT 3984

Qy   4150 CAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAG 4209
          ||| ||| |   || || || ||||||||||| |||||||  |||||||  |  ||| |
Db   3985 CAGCTGGCAAAACTTCAGGCGTGGCGCAGCGTTTACAGCAAAGTGGCGAGCTACATTGAA 4044

Qy   4210 GACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTT 4269
          || || |||||||||||||| || || |||||||| ||||||||||||||| ||||| |||
Db   4045 GATGAGCATCTGCGTCAGGCCTTCTCTTTCCACTCACTGCTGGTGGGCGGAAATCCGTTT 4104

Qy   4270 GCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTT 4329
          || || || || |||||||||| | |||||||||||||| ||||||||||||||| |||||
Db   4105 GCCACTTCCTCAATCTATACCCTGATTCATGCGCTGGAACGTGAATGGGGCGTCTGGTTC 4164

Qy   4330 CCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGC 4389
          |||||||| |||| ||||||||||||||||||||||| |  |||||| || ||  |||||
Db   4165 CCGCGCGGTGGCACGGGCGCGCTGGTGCAGGGCATGGTGAAACTGTTTGAAGATCTGGGC 4224

Qy   4390 GGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGC 4449
          |||||  ||   || |||||||   ||   || |||||| |||   |  ||  | |||| |
Db   4225 GGCGAAGTGGAGCTCAATGCCAGCGTTGCCCGGCTGGAGACCCAGGAAAACAGAATTACC 4284

Qy   4450 GGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGAC 4509
          |  || ||  |  | |  || ||   ||||    ||  ||| || |||||||| || ||
Db   4285 GCGGTGCACCTGAAAGATGGCCGGGTCTTCCCGACCCGCGCGGTTGCCTCCAACGCAGAT 4344

Qy   4510 GTGGTGCATACCTACGACAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACA 4569
          ||||| || |||||||  | |||||||  |||| |||||     |   |   |  |    |
Db   4345 GTGGTTCACACCTACCGCGAACTGCTGAGCCAGCATCCCGCTTCGCAGGCGCAGGGACGA 4404

Qy   4570 TCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAG 4629
          || ||| |    || ||||||||||||||| ||||||||  ||||||||||||||||||| 
Db   4405 TCACTGCAAAACAAACGCATGAGCAACTCACTGTTTGTGATCTATTTTGGCCTGAATCAT 4464
```

```
Qy   4630 CCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATC 4689
          |  ||  | ||||| ||||||||||| ||||| ||||| ||||||| ||||||||||||||
Db   4465 CATCACAATCAGCTGGCGCACCACACGGTCTGCTTTGGTCCGCGCTATCGTGAGTTGATT 4524

Qy   4690 GATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCC 4749
          |||||||| || ||||        |||||||| || || || || || ||||| ||||||
Db   4525 GATGAGATCTTTAACAAAGATGGCCTGGCAGAGGACTTCTCGCTCTATCTGCATGCGCCC 4584

Qy   4750 TGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCG 4809
          |||   | |||||| || ||||| |||   |||||||||||||||  || || | | |||||
Db   4585 TGCGTGACCGATCCCTCACTGGCGCCGGAGGGCTGCGGCAGCTACTACGTGCTGGCGCCA 4644

Qy   4810 GTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGAT 4869
          || ||||| ||||||||||| || ||||||||||      ||||| ||||||| ||||||||
Db   4645 GTTCCGCACCTCGGCACCGCCGATATCGACTGGGCCGTTGAAGGTCCGCGCCTGCGCGAT 4704

Qy   4870 CGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACA 4929
          || |||||||  |||||||| |||||||| ||||||||  | |||   || || || ||
Db   4705 CGCATTTTTGACTATCTGGAACAGCACTATATGCCGGGCCTGCGTAGCCAGTTGGTCACG 4764

Qy   4930 CACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCG 4989
          ||  | || || |||||||||||||| |||||   |||| ||||  |||| || ||||| 
Db   4765 CATCGCATCTTCACGCCGTTTGATTTCCGCGATGAGCTGAATGCGTATCAGGGTTCGGCC 4824

Qy   4990 TTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCC 5049
          || ||| ||||||||||  |||||||||||||||||||||| || || ||||||||
Db   4825 TTCTCGGTGGAGCCGATCCTGACGCAAAGCGCCTGGTTCCGGCCTCACAACCGCGATAAA 4884

Qy   5050 GATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGC 5109
           |||| |  |||||||||||||| ||||| || |||||||| ||||||||||  | || |||
Db   4885 AATATTAATAATCTCTATCTGGTCGGTGCAGGCACGCATCCTGGCGCGGGTATTCCAGGC 4944

Qy   5110 GTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATC 5169
          || || || ||||||||||| |||||  |  ||||||||||||||||  ||    |||||
Db   4945 GTAATTGGCTCGGCCAAGGCTACCGCAGGATTGATGCTGGAGGATCTGGC---TTGAATA 5001

Qy   5170 GACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCA 5229
          | | | |  | || || || |      | |||||||| |||||| ||||| || || ||||
Db   5002 GTCCGTCACTGCTCGATCATGCCGTAGACACCATGGAGGTGGGATCGAAAAGCTTTGCCA 5061

Qy   5230 CCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGT 5289
          ||||  | || ||||||| |||    || || ||||||  ||||||||| || || ||||
Db   5062 CCGCGTCAAAACTGTTTGGTGCCAAAACCCGACGCAGCGTGCTGATGCTCTACGCCTGGT 5121

Qy   5290 GTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATG 5349
          | ||||||||| |||||||||||||| |  ||    |||||     ||  |     ||
Db   5122 GCCGTCACTGTGATGATGTGATTGACGATCAGGTACTGGGATTCAGCAACGATACGCCAT 5181

Qy   5350 CCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACA 5409
          |  |  ||  |||  |    ||  ||  ||| ||| | ||  |||| || |  |||||
Db   5182 CGCTGCAATCCGCCGAACAGCGCCTGGCGCAGCTGGAGATGAAAACGCGTCAGGCCTATG 5241

Qy   5410 GCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACC 5469
           |||  | || ||| |||| || || |||   || |||||||| ||||| ||    |||
Db   5242 CCGGATCCCAGATGCATGAGCCCGCCTTTGCAGCCTTTCAGGAGGTGGCAATGGCACACG 5301

Qy   5470 AGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACG 5529
          |  | | ||        || ||||||||||||||||  ||||| || ||||| |  |   |
Db   5302 ATATTCTGCCTGCTTACGCTTTTGATCATCTGGCGGGCTTTGCGATGGACGTGCATGAGA 5361

Qy   5530 AACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCG 5589
           ||  ||   ||   | |||||| |||||||||||||| || |||||||||||||||| |
Db   5362 CACGCTATCAGACGCTGGATGATACGCTGCGTTACTGTTACCACGTCGCGGGCGTGGTTG 5421

Qy   5590 GTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCG 5649
          |  ||||||||||||   |  |||||||| |||||| | ||   ||| |||| ||||||||
Db   5422 GCCTGATGATGGCGCAGATTATGGGCGTACGCGACAACGCCACGCTGGATCGCGCCTGCG 5481

Qy   5650 ATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAA 5709
          || | || |||||||| ||||| || || |||||||||||| || || |||||||| ||
Db   5482 ATCTCGGTCTGGCGTTTCAGCTGACCAATATTGCGCGCGATATCGTTGAAGATGCTGACG 5541
```

```
Qy          5710 ATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGC 5769
                   || ||||||||||||||     | ||||| | | | | |||  |   |   || |  |
Db          5542 CGGGACGCTGCTATCTGCCCGCTACGTGGCTGGCTGAAGAGGGGCTTACCCGAGAGAATC 5601

Qy          5770 TGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGG 5829
                 |  | |  ||||||| ||||    ||  |   |    | ||    ||  | ||||    ||
Db          5602 TCGCCGATCCGCAAAATCGCCAGGCATTAAGCCGCGTCGCCCGCCGGCTGGTGGAAACGG 5661

Qy          5830 CGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGT 5889
                 | || ||||||||||  ||||||    | ||  | |||||||||||||||| || ||||
Db          5662 CAGAGCCCTATTATCGATCGGCGTCGGCTGGCCTGCCGGGTTTACCGCTGCGTTCAGCGT 5721

Qy          5890 GGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCG 5949
                 |||| || ||||| || |     || || ||  |||| ||    || ||    || || |
Db          5722 GGGCGATTGCTACCGCGCAGCAGGTCTACCGTAAAATCGGTATGAAGGTGGTTCAGGCGG 5781

Qy          5950 GTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGG 6009
                  |   || || ||||||  ||| |||  ||||||| | |   || |||||||| |||||||
Db          5782 CTTCACAGGCGTGGGATCAACGCCAGTCCACCAGCACACCAGAGAAACTGGCACTGCTGG 5841

Qy          6010 TGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTG 6069
                 ||   |   | |||  |||| | ||||| || |||   ||||  |   | ||  | ||||
Db          5842 TGGCGGCATCGGGTCAGGCGGTTACTTCCCGGGTGGCGCGTCACGCTCCACGCTCCGCTG 5901

Qy          6070 GTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAG 6129
                  ||| |||||||| ||         ||||    || ||||| |  ||||||||| || ||||
Db          5902 ATCTCTGGCAGCGCCC------CGTTTAACGCCATGACGTTCTCGCAGCGTCGCCTGCAA 5955

Qy          6130 CTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCGCAC 6189
                  ||   |||||| |||||||| || || ||||||||    ||||| ||     || || ||
Db          5956 TTTTGACAGCGGCGGCGCGTAAAGAAAGCCAAACGAGACACAGCCCTCTTTGCCTCGTAC 6015

Qy          6190 CGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGATATA 6249
                 |||||||||||||||||||||||| || |  |     | ||| ||   |||||| | |||
Db          6016 CGCATGATGCATGCGGTGCGCCATATAAAGCCTGCGTAAATAACCCCGGCGCGGAACATA 6075

Qy          6250 GCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCCGTA 6309
                 |||||| |||||||| || ||||||||||| || ||||| || || ||||||  ||||||
Db          6076 GCGGAATGGCCAGCGCTGGTGCACCAGGCCGTCATGCACAATAAAATAGAGCAGGCCGTA 6135

Qy          6310 CGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAAT 6369
                    |||||| || || || ||||||||||  ||||| |  ||       || | ||||||
Db          6136 GAGCGTCATACCCGCGCCGATCCACTGCAAGGGCCAGACACCCGTGCTGCCCAGATAAAT 6195

Qy          6370 CAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACC 6429
                  |||| ||||| |||  | || ||||| ||  ||||| || || || | |||||   ||
Db          6196 GAGCAAAATCGACAGAGCGGCGAACACTACGGCATAGAGGTCATTAACCTCAAACCAGCC 6255

Qy          6430 GCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTT 6489
                   | ||||||||||| || |||||||||||||||||| ||||| || ||||| |||||| |
Db          6256 TTTATGCGGTTCATGATGTGACAGATGCCAGCCCCAACCCCAGCCATGCATAATGTATCT 6315

Qy          6490 ATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTT 6549
                  |||| ||| || ||    || ||||| |  |  |  ||   ||| ||  || ||    ||
Db          6316 GTGCGCCAGTGCAGCCGTTATCTCCATTCCGATAACAGTGACCAGAACAATCAGGGCATT 6375

Qy          6550 CCATAACCAGAGCATTGTTCGTCC 6573
                 ||| | ||| | ||| |||  |||
Db          6376 CCAAATCCACAACATAGTTTCTCC 6399


RESULT 2
ERWCRT
LOCUS       ERWCRT                  6918 bp    DNA     linear   BCT 24-AUG-2002
DEFINITION  Erwinia uredovora DNA carotenoid biosynthesis gene cluster (crtE,
            crtX, crtY, crtI, crtB, crtZ), complete cds.
ACCESSION   D90087
VERSION     D90087.2  GI:22474502
KEYWORDS    .
SOURCE      Pantoea ananatis
  ORGANISM  Pantoea ananatis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1  (bases 1 to 6918)
  AUTHORS   Misawa,N., Nakagawa,M., Kobayashi,K., Yamano,S., Izawa,Y.,
            Nakamura,K. and Harashima,K.
  TITLE     Elucidation of the Erwinia uredovora carotenoid biosynthetic
            pathway by functional analysis of gene products expressed in
            Escherichia coli
  JOURNAL   J. Bacteriol. 172 (12), 6704-6712 (1990)
  MEDLINE   91072214
   PUBMED   2254247
COMMENT     On Aug 23, 2002 this sequence version replaced gi:216681.
            These data kindly submitted in computer readable form by:  Norihiko
            Misawa
            Marine Biotechnology Institute
            3-75-1, Heita, Kamaishi-shi,
            Iwate 026-0001
            Japan
            e-mail: norihiko.misawa@mbio.jp
            Phone:  81-193-26-6581
            Fax:    81-193-26-6584.
FEATURES             Location/Qualifiers
     source          1. .6918
                     /organism="Pantoea ananatis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:553"
                     /note="carotenoid biosynthesis genes"
     gene            225. .1133
                     /gene="crtE"
     CDS             225. .1133
                     /gene="crtE"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA14124.1"
                     /db_xref="GI:216682"
                     /translation="MTVCAKKHVHLTRDAAEQLLADIDRRLDQLLPVEGERDVVGAAM
                     REGALAPGKRIRPMLLLLTARDLGCAVSHDGLLDLACAVEMVHAASLILDDMPCMDDA
                     KLRRGRPTIHSHYGEHVAILAAVALLSKAFGVIADADGLTPLAKNRAVSELSNAIGMQ
                     GLVQGQFKDLSEGDKPRSAEAILMTNHFKTSTLFCASMQMASIVANASSEARDCLHRF
                     SLDLGQAFQLLDDLTDGMTDTGKDSNQDAGKSTLVNLLGPRAVEERLRQHLQLASEHL
                     SAACQHGHATQHFIQAWFDKKLAAVS"
     gene            1143. .2438
                     /gene="crtX"
     CDS             1143. .2438
                     /gene="crtX"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA14125.1"
                     /db_xref="GI:216683"
                     /translation="MSHFAAIAPPFYSHVRALQNLAQELVARGHRVTFIQQYDIKHLI
                     DSETIGFHSVGTDSHPPGALTRVLHLAAHPLGPSMLKLINEMARTTDMLCRELPQAFN
                     DLAVDGVIVDQMEPAGALVAEALGLPFISVACALPLNREPDMPLAVMPFEYGTSDAAR
                     ERYAASEKIYDWLMRRHDRVIAEHSHRMGLAPRQKLHQCFSPLAQISQLVPELDFPRK
                     ALPACFHAVGPLRETHAPSTSSSRYFTSSEKPRIFASLGTLQGHRYGLFKTIVKACEE
                     IDGQLLLAHCGRLTDSQCEELARSRHTQVVDFADQSAALSQAQLAITHGGMNTVLDAI
                     NYRTPLLALPLAFDQPGVASRIVYHGIGKRASRFTTSHALARQMRSLLTNVDFQQRMA
                     KIQTALRLAGGTMAAADIIEQVMCTGQPVLSGSGYATAL"
     gene            2422. .3570
                     /gene="crtY"
     CDS             2422. .3570
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA14126.1"
                     /db_xref="GI:216684"
                     /translation="MQPHYDLILVGAGLANGLIALRLQQQQPDMRILLIDAAPQAGGN
                     HTWSFHHDDLTESQHRWIAPLVVHHWPDYQVRFPTRRRKLNSGYFCITSQRFAEVLQR
                     QFGPHLWMDTAVAEVNAESVRLKKGQVIGARAVIDGRGYAANSALSVGFQAFIGQEWR
                     LSHPHGLSSPIIMDATVDQQNGYRFVYSLPLSPTRLLIEDTHYIDNATLDPECARQNI
                     CDYAAQQGWQLQTLLREEQGALPITLSGNADAFWQQRPLACSGLRAGLFHPTTGYSLP
                     LAVAVADRLSALDVFTSASIHHAITHFARERWQQQGFFRMLNRMLFLAGPADSRWRVM
                     QRFYGLPEDLIARFYAGKLTLTDRLRILSGKPPVPVLAALQAIMTTHR"
     gene            3582. .5060
                     /gene="crtI"
     CDS             3582. .5060
```

```
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA14127.1"
                     /db_xref="GI:216685"
                     /translation="MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYE
                     DQGFTFDAGPTVITDPSAIEELFALAGKQLKEYVELLPVTPFYRLCWESGKVFNYDND
                     QTRLEAQIQQFNPRDVEGYRQFLDYSRAVFKEGYLKLGTVPFLSFRDMLRAAPQLAKL
                     QAWRSVYSKVASYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLIHALEREWGVWFPRG
                     GTGALVQGMIKLFQDLGGEVVLNARVSHMETTGNKIEAVHLEDGRRFLTQAVASNADV
                     VHTYRDLLSQHPAAVKQSNKLQTKRMSNSLFVLYFGLNHHHDQLAHHTVCFGPRYREL
                     IDEIFNHDGLAEDFSLYLHAPCVTDSSLAPEGCGSYYVLAPVPHLGTANLDWTVEGPK
                     LRDRIFAYLEQHYMPGLRSQLVTHRMFTPFDFRDQLNAYHGSAFSVEPVLTQSAWFRP
                     HNRDKTITNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI"
     gene            5057. .5986
                     /gene="crtB"
     CDS             5057. .5986
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA14128.2"
                     /db_xref="GI:22474503"
                     /translation="MNNPSLLNHAVETMAVGSKSFATASKLFDAKTRRSVLMLYAWCR
                     HCDDVIDDQTLGFQARQPALQTPEQRLMQLEMKTRQAYAGSQMHEPAFAAFQEVAMAH
                     DIAPAYAFDHLEGFAMDVREAQYSQLDDTLRYCYHVAGVVGLMMAQIMGVRDNATLDR
                     ACDLGLAFQLTNIARDIVDDAHAGRCYLPASWLEHEGLNKENYAAPENRQALSRIARR
                     LVQEAEPYYLSATAGLAGLPLRSAWAIATAKQVYRKIGVKVEQAGQQAWDQRQSTTTP
                     EKLTLLLAASGQALTSRMRAHPPRPAHLWQRPL"
     gene            complement(5925. .6452)
                     /gene="crtZ"
     CDS             complement(5925. .6452)
                     /gene="crtZ"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="BAA14129.1"
                     /db_xref="GI:216687"
                     /translation="MLWIWNALIVFVTVIGMEVIAALAHKYIMHGWGWGWHLSHHEPR
                     KGAFEVNDLYAVVFAALSILLIYLGSTGMWPLQWIGAGMTAYGLLYFMVHDGLVHQRW
                     PFRYIPRKGYLKRLYMAHRMHHAVRGKEGCVSFGFLYAPPLSKLQATLRERHGARAGA
                     ARDAQGGEDEPASGK"
ORIGIN

  Query Match             37.5%;  Score 2626.2;  DB 1;  Length 6918;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4008;  Conservative    0;  Mismatches 2168;  Indels   21;  Gaps    6;

Qy          400 CTTGAACATTTACTGCCTGCCGGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCC 459
                ||||| || ||| |||| |  |      || | |||   ||| |||||||||||||||
Db          303 CTTGATCAGTTATTGCCCGTGGAGGGAGAACGGGATGTTGTGGGTGCCGCGATGCGTGAA 362

Qy          460 GGAACGCTGGCGCAGGGCAAACGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGAT 519
                ||  ||||||| | ||| ||||||||||| ||  | || ||| ||||| | || ||||||
Db          363 GGTGCGCTGGCACCGGGAAAACGTATTCGCCCCATGTTGCTGTTGCTGACCGCCCGCGAT 422

Qy          520 ATGGGTTGCGAGCTGACGCAAAATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTG 579
                 |||||||||   | |  ||  | ||  | || ||| | |||||||| |||||||||||
Db          423 CTGGGTTGCGCTGTCAGCCATGACGGATTACTGGATTTGGCCTGTGCGGTGGAAATGGTC 482

Qy          580 CACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGT 639
                |||||||| |||||||| || || || || ||||  |||||  | ||| || |||| ||
Db          483 CACGCGGCTTCGCTGATCCTTGACGATATGCCCTGCATGGACGATGCGAAGCTGCGGCGC 542

Qy          640 GGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCG 699
                || ||||||||| | |||    | |  || ||  | ||||| || || || ||  | ||
Db          543 GGACGCCCTACCATTCATTCTCATTACGGAGAGCATGTGGCAATACTGGCGGCGGTTGCC 602

Qy          700 CTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAA 759
                 |||| ||    || ||||  || ||||||  ||||   ||  |  |  |  |    |||
Db          603 TTGCTGAGTAAAGCCTTTGGCGTAATTGCCGATGCAGATGGCCTCACGCCGCTGGCAAAA 662

Qy          760 TCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAA 819
                  |   ||| || ||||||| ||    ||| ||||| |||| || || || || || ||
Db          663 AATCGGGCGGTTTCTGAACTGTCAAACGCCATCGGCATGCAAGGATTGGTTCAGGGTCAG 722

Qy          820 TTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAAC 879
                ||| ||||||||   || ||     ||   |||||| | ||||| ||    ||||| ||
Db          723 TTCAAGGATCTGTCTGAAGGGGATAAGCCGCGCAGCGCTGAAGCTATTTTGATGACGAAT 782

Qy          880 GAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGAC 939
                 |  | |||||||||  ||||||| | ||| |  |||| |||||  ||||||  ||  |
Db          783 CACTTTAAAACCAGCACGCTGTTTTGTGCCTCCATGCAGATGGCCTCGATTGTTGCGAAT 842

Qy          940 GCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTT 999
                || ||     | | |||  |  | ||       ||  | |  ||| | || ||||| |||
Db          843 GCCTCCAGCGAAGCGCGTGATTGCCTGCATCGTTTTTCACTTGATCTTGGTCAGGCATTT 902

Qy         1000 CAACTGCTCGACGACCTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGAT 1059
                |||||||| |||||  |  |||| ||    |   |||||||||| |||    | |||||
Db          903 CAACTGCTGGACGATTTGACCGATGGCATGACCGACACCGGTAAGGATAGCAATCAGGAC 962

Qy         1060 CAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGC 1119
                   || ||||| ||||||||  |  || | ||  |    |  |  |||   || ||| |
Db          963 GCCGGTAAATCGACGCTGGTCAATCTGTTAGGCCCGAGGGCGGTTGAAGAACGTCTGAGA 1022

Qy         1120 GATCACCTGCGCAGCGCAGATGCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACT 1179
                 | || || |     ||   ||  || ||  |    |||||||| | |||   |||||||
Db         1023 CAACATCTTCAGCTTGCCAGTGAGCATCTCTCTGCGGCCTGCCAACACGGGCACGCCACT 1082

Qy         1180 CGCCAATATATGCACGCGCTGTTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTC 1239
                |  || | ||| || ||   |||| |  || | || ||     |||  | ||   |   |
Db         1083 CAACATTTTATTCAGGCCTGGTTTGACAAAAAACTCGCTGCCGTCAGTTAAGGATGCTGC 1142

Qy         1240 AGCCGGTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGC 1299
                |     || |||| || |||| |||||| |||||  | ||||||||  ||| ||| || |
Db         1143 A-----TGAGCCATTTCGCGGCGATCGCACCGCCTTTTTACAGCCATGTTCGCGCATTAC 1197

Qy         1300 AGGCGTTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCG 1359
                ||    | || ||    ||| | |||||||| |||||  | || || || ||||||  ||
Db         1198 AGAATCTCGCTCAGGAACTGGTCGCGCGCGGTCATCGGGTGACCTTTATTCAGCAATACG 1257

Qy         1360 ATGCCCGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGC 1419
                ||         ||| |      ||||  ||| |  |||  | |||||||   | | |  |
Db         1258 ATATTAAACACTTGATCGATAGCGAAACCATTGGATTTCATTCCGTCGGGACAGACAGCC 1317

Qy         1420 ATCCTGCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGC 1479
                ||||  ||||  || |  ||| |||| | || | |  |||    ||    || | |||
Db         1318 ATCCCCCCGGCGCGTTAACGCGCGTGCTACACCTGGCGGCTCATCCTCTGGGGCCGTCAA 1377

Qy         1480 TGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTG 1539
                || |     | ||| | ||  | ||| |   |||||||||||||||||||||||| ||
Db         1378 TGCTGAAGCTCATCAATGAAATGGCGCGCACCACCGATATGCTGTGCCGCGAACTCCCCC 1437

Qy         1540 CGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCG 1599
                 || | | || |   ||  | |||||||||| || |  ||  |||||||| |||| ||||
Db         1438 AGGCATTTAACGATCTGGCCGTCGATGGCGTCATTGTTGATCAAATGGAACCGGCAGGCG 1497

Qy         1600 GATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCA 1659
                   | || |||||||| |||   ||||||||| | || || ||||||||  ||||  |||
Db         1498 CGCTCGTTGCTGAAGCACTGGGACTGCCGTTTATCTCTGTCGCCTGCGCGCTGCCTCTCA 1557

Qy         1660 ATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAG 1719
                |||||||| | |  || || || ||||| ||||| |||   |  |        |||   |
Db         1558 ATCGTGAACCGGATATGCCCCTGGCGGTTATGCCTTTCGAATACGGGACCAGCGACGCGG 1617

Qy         1720 CGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCG 1779
                | |   |||||| |   |||||     | || |||||  |  | ||||||||||| | |
Db         1618 CTCGCGAACGTTATGCCGCCAGTGAAAAAATTTATGACTGGCTAATGCGTCGTCATGACC 1677

Qy         1780 ACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGT 1839
                  || ||   | |||||   |      |    ||  |    ||||      | || ||||
Db         1678 GTGTCATTGCCGAACACAGCCACAGAATGGGCTTAGCCCCCCGGCAAAAGCTTCACCAGT 1737

Qy         1840 GCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGC 1899
                |  | ||||| ||||| |||||||||||| | || || |   | |||||||| ||  |
Db         1738 GTTTTTCGCCACTGGCGCAAATCAGCCAGCTTGTTCCTGAACTGGATTTTCCCCGCAAAG 1797
```

```
Qy    1900 AACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGC 1959
              | || || || | ||| ||||| |||| | || |||| ||    | ||  |  ||  |  ||
Db    1798 CGTTACCGGCTTGTTTTCATGCCGTCGGGCCTCTGCGCGAAACGCACGCACCGTCAACGT 1857

Qy    1960 TCCATGCGCCCTG---GCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGC 2016
                 | |  |      || |        | ||   | |||  |||||||||||| ||||
Db    1858 CTTCATCCCGTTATTTTACATCCTCAGAAAAACCCCGGATTTTCGCCTCGCTGGGCACGC 1917

Qy    2017 TGCAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGC 2076
           | || || || || |   ||||||||       | | | | || ||     |  |
Db    1918 TTCAGGGACACCGTTATGGGCTGTTTAAAACGATAGTGAAAGCCTGTGAAGAAATTGACG 1977

Qy    2077 TATCGCTGGTGATCGCCCATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGC 2136
               |||  || | ||||| |||||  |  | |  | |   |||    |  |||||| ||
Db    1978 GTCAGCTCCTGTTAGCCCACTGTGGTCGTCTTACGGACTCTCAGTGTGAAGAGCTGGCGC 2037

Qy    2137 TCGCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGC 2196
                 |     |   ||||  |||||| | |||||||   || || ||    || ||||
Db    2038 GAAGCCGTCATACACAGGTGGTGGATTTTGCCGATCAGTCAGCCGCGCTGTCTCAGGCGC 2097

Qy    2197 AGCTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGC 2256
           |||||   ||||| || | |||  | || |  |  ||||| ||  |  | | | | ||||
Db    2098 AGCTGGCGATCACCCACGGCGGCATGAATACGGTACTGGACGCGATTAATTACCGGACGC 2157

Qy    2257 CGATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGC 2316
           |  |  | ||||| ||| | || |||||||||||||||||| ||| | ||||| |  |  |
Db    2158 CCCTTTTAGCGCTTCCGCTGGCCTTTGATCAGCCCGGCGTCGCGTCACGCATCGTTTATC 2217

Qy    2317 ATGACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGC 2376
           | | | | ||    || || || ||||||| |      |||    |||  |  ||  |||
Db    2218 ACGGCATCGGCAAGCGTGCTTCCCGCTTTACCACCAGCCATGCTTTGGCTCGTCAGATGC 2277

Qy    2377 AACAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGC 2436
                 |||||||| |||     |    ||  || ||| |    || ||| |    || |
Db    2278 GTTCATTGCTGACCAACGTCGACTTTCAGCAGCGCATGGCGAAAATCCAGACAGCCCTTC 2337

Qy    2437 AGCGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGC 2496
                ||||| ||   |  |   || ||||| ||| | |||||||   |||| || | |
Db    2338 GTTTGGCAGGGGGCACCATGGCCGCTGCCGATATCATTGAGCAGGTTATGTG-CACCGGT 2396

Qy    2497 AAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGC 2556
            || |   ||   |  || |     ||||   |||| || ||| ||||| | || || ||
Db    2397 CAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCGTGGGGGC 2456

Qy    2557 TGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATG 2616
           |||||| ||||||||| | || || ||||||||| |  ||   ||| ||  |  ||
Db    2457 TGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATATGCGTAT 2516

Qy    2617 CCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAG 2676
             || |  | ||   ||   | ||  |||  || |||||||||| ||||| |||| |||
Db    2517 TTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTCACCACGA 2576

Qy    2677 CGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGG 2736
            ||| | ||     |||| ||| ||| |    |||||| |       |  ||| | |
Db    2577 TGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACTGGCCCGA 2636

Qy    2737 TTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGC 2796
            |||||||| || |||||  | |  || || || ||| || | |  || | ||  ||  |
Db    2637 CTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTTGTATTAC 2696

Qy    2797 ATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACAC 2856
            ||      ||| ||      | |  |      | | ||     |  ||||||   | ||
Db    2697 TTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGATGGATAC 2756

Qy    2857 AGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGC 2916
            || ||   | |||| ||  |       ||   | ||    | || |  |   | | |||
Db    2757 CGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTATCGGTGC 2816

Qy    2917 GCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCA 2976
            |  | |||||||| || || ||      | | |   ||   |||    |||| |  ||
Db    2817 CCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGGGCTTCCA 2876

Qy    2977 GGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCT 3036
           || |||| |||| || || ||||   ||   || ||||| ||  |   |   || ||  |
Db    2877 GGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTCCCATTAT 2936

Qy    3037 GATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAG 3096
            ||||||||||| |||||||||||||   |||||||| || || ||||  |||||||||
Db    2937 CATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGCCGCTCTC 2996

Qy    3097 CGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAA 3156
             |     |  | || |||||||| || ||  || | ||  |  |  |  |  | |      |
Db    2997 GCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAGATCCTGA 3056

Qy    3157 CACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCT 3216
              ||| || ||  | ||  |||||||||||   || || || |||  ||    || ||
Db    3057 ATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTCAGACACT 3116

Qy    3217 GCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTG 3276
           |||||| ||||| || |||   ||||| ||||||| |||   |||||   ||   ||||||
Db    3117 GCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACGCATTCTG 3176

Qy    3277 GCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCAC 3336
           |   ||||||||||  ||  || || || || ||  | || ||||| ||||| || | |||||
Db    3177 G---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATCCTACCAC 3233

Qy    3337 CGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCAC 3396
           ||| || || |||||   || || ||  | || ||    |    || |      |
Db    3234 CGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTGATGTCTT 3293

Qy    3397 CGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACA 3456
                |   | |  |   |||    ||   | ||||||||  |||  ||||   | ||
Db    3294 TACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGCAGCAGCA 3353

Qy    3457 GCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTG 3516
           | | ||||||||  |||| || |||||||||||||||| |||||     |  |    |||||
Db    3354 GGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATTCACGCTG 3413

Qy    3517 GCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGG 3576
           ||| || ||||| ||||||||  |  |  |   ||  ||||||  ||| ||||| || ||
Db    3414 GCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTTATGCGGG 3473

Qy    3577 GCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGG 3636
             ||||    |||  ||||      || ||||| |||||||||||||| | ||  |
Db    3474 AAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTCCGGTATT 3533

Qy    3637 TGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAAATGAAACGCACT 3696
            | |||  ||  |  ||  |  |||   ||        | |   ||  | ||||||  |||
Db    3534 AGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAAACCAACT 3593

Qy    3697 TATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGC 3756
              || |||||| |||||||| || |||||||| ||||| |||| || ||||| |||||
Db    3594 ACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGCTGCGGGG 3653

Qy    3757 ATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAG 3816
           || ||   |    | || || || || || ||||| |||||| || || || ||| |  |||
Db    3654 ATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGTCTACGAG 3713

Qy    3817 GACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGAA 3876
           ||    || |||||||| ||||| || || ||||| |||||||||||||| ||||| |||
Db    3714 GATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGCCATTGAA 3773

Qy    3877 GAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACG 3936
           ||  ||||  | ||||||||||||  | | |  || || |||| ||| ||||||| |||
Db    3774 GAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCCGGTTACG 3833

Qy    3937 CCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCCG 3996
           || || || |||||||||| ||||||    || ||     || ||||||| ||  ||||  |
Db    3834 CCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGATCAAACC 3893

Qy    3997 CTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTT 4056
           | ||| ||   ||||||      ||| |||   || |  ||||| ||||| |||||| |||
Db    3894 CGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCGTCAGTTT 3953

Qy    4057 CTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTT 4116
           || | |||||     |  || |||| ||| || ||||| || ||||| || || || |||
Db   3954 CTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGTCCCTTTT 4013

Qy   4117 CTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGC 4176
           |   | |  | |||||||| |||| ||| || ||  |||     ||||| |||||| |
Db   4014 TTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGCATGGAGA 4073

Qy   4177 AGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCC 4236
          ||||| ||||| | ||| || |  |  ||  | || || |||||||| ||||||||||
Db   4074 AGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGCGTTTTCT 4133

Qy   4237 TTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAATT 4296
          ||||||||  || |||||||||| ||||| || || || ||||| || ||||| || ||
Db   4134 TTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATACGTTGATA 4193

Qy   4297 CATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTG 4356
          || |||||||||||||| |||||||| |||||||||||| ||||||||||||||  | ||
Db   4194 CACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGCATTAGTT 4253

Qy   4357 CAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTG 4416
          ||||| |||      |||||  ||||  |||| |||||  |  |  | || |||  |||
Db   4254 CAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGCCAGAGTC 4313

Qy   4417 AGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGC 4476
          |||||  ||||||| |  || ||    |||   | ||| || ||||||| ||| ||  |
Db   4314 AGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGGTCGCAGG 4373

Qy   4477 TTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCTT 4536
          |||    |    || || || || ||||| || ||||| |||||||||   | | ||| |
Db   4374 TTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGACCTGTTA 4433

Qy   4537 CGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAAC 4596
           |||| || ||    ||  | || | |   | |     ||| ||  ||||||||||||| |||
Db   4434 AGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCATGAGTAAC 4493

Qy   4597 TCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACACC 4656
          || ||||||||| |||||||||||  ||||||| |  |||||| ||||||||||| |||||
Db   4494 TCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCATCACACG 4553

Qy   4657 GTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTG 4716
          || ||||| ||||||||||| || ||| ||| || || ||||| ||          ||
Db   4554 GTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGATGGCCTC 4613

Qy   4717 GCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCG 4776
          ||||| || || |||||||| |||||||||||||||   ||  ||| |||| ||||| ||
Db   4614 GCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACTGGCGCCT 4673

Qy   4777 CCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATC 4836
             || |||||||| |  |||||||| ||||||||||||||||| | |||||||||  || ||
Db   4674 GAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGCGAACCTC 4733

Qy   4837 GACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCAC 4896
          ||||||      || || ||     | |||||| || |||||||| || || ||||||||
Db   4734 GACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGAGCAGCAT 4793

Qy   4897 TACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTTT 4956
          |||||||| || |||||    ||  | || || ||| | ||||||||||||||||||||||
Db   4794 TACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTTTGATTTT 4853

Qy   4957 CGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAA 5016
          ||||||  |||  ||||| |||| ||||| || |||||  |||||||  || | || ||
Db   4854 CGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCTTACCCAG 4913

Qy   5017 AGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGT 5076
          ||||||||||| || |||||||||||||||||      || |  |||||||| ||||| ||
Db   4914 AGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCTGGTCGGC 4973

Qy   5077 GCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCC 5136
          || || |||||||| ||||| ||| | || ||||| |||||| ||||| || || || ||
Db   4974 GCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGCGACAGCA 5033

Qy   5137 AGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACGC 5196
           |  |||||||||||||||       ||||||   | | | |||||  |  || |
Db   5034 GGTTTGATGCTGGAGGATC---TGATATGAATAATCCGTCGTTACTCAATCATGCGGTCG 5090

Qy   5197 AAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGA 5256
          |||| ||||| || |||||||| ||||| || || ||| | ||| | |||||||||   |
Db   5091 AAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGATGCAAAAA 5150

Qy   5257 CGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGATG 5316
          | || ||||||   |||||||| || || || |||| || || || || ||||| |
Db   5151 CCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTTATTGACG 5210

Qy   5317 GGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGC 5376
            || |||||||||      | |  ||||| |||| |  ||    |  |   |||| ||
Db   5211 ATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAACGTCTGA 5270

Qy   5377 AGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCGT 5436
           ||| ||  | ||  |||| ||||  |||||    ||  ||| | ||| ||| | ||||||||||
Db   5271 TGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAACCGGCGT 5330

Qy   5437 TTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGATC 5496
          ||  ||| |||||||||||||| ||   |||  |  |  |  |  |         ||||||||||
Db   5331 TTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCGTTTGATC 5390

Qy   5497 ATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACGC 5556
          |||||||||||||||| ||||||| |||| | |  || ||||       |  ||||| ||||
Db   5391 ATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGC 5450

Qy   5557 TGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCG 5616
          |||| || ||||||||| || ||||| || |||||| ||||||||||||||   | | |||||||
Db   5451 TGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATCATGGGCG 5510

Qy   5617 TGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTA 5676
          |||| ||  | ||   |||   || | ||||| ||  | ||   | ||||| || ||| | || |
Db   5511 TGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAGTTGACCA 5570

Qy   5677 ACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCT 5736
          | ||||| |||| ||||| || |||||  |    || ||||||| ||||||||||  |  ||
Db   5571 ATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCGGCAAGCT 5630

Qy   5737 GGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCGC 5796
          |||| || || |  ||  |   |   || |      | || ||||| || | |||   ||||
Db   5631 GGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGTCAGGCGC 5690

Qy   5797 TCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGAT 5856
          |   |    | ||    ||||| |||  || || |||||  || |||   || ||   |
Db   5691 TGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCTGCCACAG 5750

Qy   5857 CCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTT 5916
          ||||  |  | || || || ||||| || || ||||| ||||||||||||       ||||
Db   5751 CCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAGCAGGTTT 5810

Qy   5917 ATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGC 5976
          | ||  |||| || ||||||||| | || |||||  ||| |||||||||   ||||||
Db   5811 ACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAGCGGCAGT 5870

Qy   5977 GCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTT 6036
            || |  |     |||||| |  |||||||| ||  |   | |||  ||||  | ||||
Db   5871 CAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCCCTTACTT 5930

Qy   6037 CGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTT 6096
          | ||  ||     || |   || || || ||   ||| |||||| ||      | |
Db   5931 CCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC------TCT 5984

Qy   6097 ACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAA 6156
          |   || || |  |  || ||||| || ||  || ||   ||| || |||| ||||||||
Db   5985 AGCGCCATGTCTTTCCCGGAGCGTCGCCTGAAGTTTTGACAGGGGCGGCGCATAGAGGAA 6044

Qy   6157 ACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGTA 6216
           ||||| || || || ||||||    || |  || || |||||||| |||||||||| ||
Db   6045 GCCAAAAGAAACACAACCTTCTTTGCCCCTGACGGCGTGATGCATACGGTGCGCCATATA 6104

Qy   6217 TAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCAG 6276
           || || || || ||||| ||||| || ||||||||| ||||| ||||||||||||||
Db   6105 CAACCGTTTGAGGTAGCCCTTGCGTGGAATATAGCGGAATGGCCAACGTTGATGCACCAG 6164
```

```
Qy        6277 GCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACTG 6336
                || |||||||||||| || |||||    || |||| |||||| || || |||||||||||
Db        6165 CCCGTCGTGCACCATAAAATAGAGTAATCCATACGCCGTCATACCTGCGCCAATCCACTG 6224

Qy        6337 CAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACAC 6396
                |||||||||||| |||        || | |||||||||||  ||||  | | | ||||| ||
Db        6225 GAGCGGCCACATTCCTGTACTGCCCAGATAAATCAGCAGGATCGATAATGCAGCAAAAAC 6284

Qy        6397 CACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGATG 6456
               ||| |||||||||||||||| |  ||||||  |||  |  |||||||||| || || |||||
Db        6285 CACGGCATAAAGATCGTTAACTTCAAACGCACCTTTACGCGGTTCATGATGTGAAAGATG 6344

Qy        6457 CCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCAT 6516
               ||| ||||| ||||| |||||||||||||||||||| || | ||| || || |  | ||||||
Db        6345 CCATCCCCAACCCCAGCCGTGCATGATGTATTTGTGTGCCAGTGCAGCAATCACTTCCAT 6404

Qy        6517 CACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATTGTTCGTCCATT 6576
                 | | ||||||  | || | ||| ||  | ||||| | ||| | |||  ||  |||  |
Db        6405 GCCAATCACGGTAACGAAAACGATCAGGGCATTCCAAATCCACAACATAATTTCTCCGGT 6464

Qy        6577 TGTGGAAAAGGGAAGTA 6593
                | |      || || |
Db        6465 AGAGACGTCTGGCAGCA 6481


RESULT 3
I12910
LOCUS       I12910                  6918 bp    DNA     linear   PAT 26-JUL-1995
DEFINITION  Sequence 13 from patent US 5429939.
ACCESSION   I12910
VERSION     I12910.1  GI:910887
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6918)
  AUTHORS   Misawa,N., Kobayashi,K., Nakamura,K. and Yamano,S.
  TITLE     DNA sequences useful for the synthesis of carotenoids
  JOURNAL   Patent: US 5429939-A 13 04-JUL-1995;
FEATURES             Location/Qualifiers
     source          1..6918
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             37.5%;  Score 2624.6;  DB 6;  Length 6918;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4007;  Conservative    0;  Mismatches 2169;  Indels   21;  Gaps    6;

Qy         400 CTTGAACATTTACTGCCTGCCGGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCC 459
               ||||| || ||| |||| |  |      || | |||   ||| |||||||||||||||||
Db         303 CTTGATCAGTTATTGCCCGTGGAGGGAGAACGGGATGTTGTGGGTGCCGCGATGCGTGAA 362

Qy         460 GGAACGCTGGCGCAGGGCAAACGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGAT 519
               ||  ||||||| | ||| ||||||||||| ||  | || ||| ||||| | || ||||||
Db         363 GGTGCGCTGGCACCGGGAAAACGTATTCGCCCCATGTTGCTGTTGCTGACCGCCCGCGAT 422

Qy         520 ATGGGTTGCGAGCTGACGCAAAATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTG 579
                |||||||||   | |  ||  | ||  | || ||| | ||||||| ||||||||||||
Db         423 CTGGGTTGCGCTGTCAGCCATGACGGATTACTGGATTTGGCCTGTGCGGTGGAAATGGTC 482

Qy         580 CACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGT 639
               |||||||| ||||||||| || || || || ||||  |||||  | ||| || |||| || 
Db         483 CACGCGGCTTCGCTGATCCTTGACGATATGCCCTGCATGGACGATGCGAAGCTGCGGCGC 542

Qy         640 GGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCG 699
               || ||||||||| | |||    | |  || ||  | |||||| || || || ||  | || 
Db         543 GGACGCCCTACCATTCATTCTCATTACGGAGAGCATGTGGCAATACTGGCGGCGGTTGCC 602

Qy         700 CTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAA 759
                |||| ||    || ||||  || ||||||  ||||   ||  |  |  |  |    |||
Db         603 TTGCTGAGTAAAGCCTTTGGCGTAATTGCCGATGCAGATGGCCTCACGCCGCTGGCAAAA 662

Qy         760 TCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAA 819
                 |   ||| || ||||||| ||    ||| ||||| |||| || || || || || || 
Db         663 AATCGGGCGGTTTCTGAACTGTCAAACGCCATCGGCATGCAAGGATTGGTTCAGGGTCAG 722

Qy         820 TTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAAC 879
               ||| ||||||||   || ||     ||   |||||| | ||||| ||    ||||| || 
Db         723 TTCAAGGATCTGTCTGAAGGGGATAAGCCGCGCAGCGCTGAAGCTATTTTGATGACGAAT 782

Qy         880 GAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGAC 939
                |  | |||||||||  ||||||| | ||| |  |||| |||||  ||||||  ||  | 
Db         783 CACTTTAAAACCAGCACGCTGTTTTGTGCCTCCATGCAGATGGCCTCGATTGTTGCGAAT 842

Qy         940 GCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTT 999
               || ||     | | |||  |  | ||       ||  | |  ||| | || ||||| |||
Db         843 GCCTCCAGCGAAGCGCGTGATTGCCTGCATCGTTTTTCACTTGATCTTGGTCAGGCATTT 902

Qy        1000 CAACTGCTCGACGACCTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGAT 1059
               |||||||| |||||  |  |||| ||    |   |||||||||| |||     | ||||| 
Db         903 CAACTGCTGGACGATTTGACCGATGGCATGACCGACACCGGTAAGGATAGCAATCAGGAC 962

Qy        1060 CAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGC 1119
                  || ||||| ||||||||  |  || | ||  |    |  |  |||   || ||| | 
Db         963 GCCGGTAAATCGACGCTGGTCAATCTGTTAGGCCCGAGGGCGGTTGAAGAACGTCTGAGA 1022

Qy        1120 GATCACCTGCGCAGCGCAGATGCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACT 1179
                | || || |     ||   ||  || ||  |    |||||||| | |||    |||||||
Db        1023 CAACATCTTCAGCTTGCCAGTGAGCATCTCTCTGCGGCCTGCCAACACGGGCACGCCACT 1082

Qy        1180 CGCCAATATATGCACGCGCTGTTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTC 1239
               |  || | ||| || ||   |||| |  || | || ||     |||  | ||   |   |
Db        1083 CAACATTTTATTCAGGCCTGGTTTGACAAAAAACTCGCTGCCGTCAGTTAAGGATGCTGC 1142

Qy        1240 AGCCGGTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGC 1299
               |     || |||| || ||||| |||||| |||||  | ||||||||  ||| ||| || |
Db        1143 A-----TGAGCCATTTCGCGGCGATCGCACCGCCTTTTTACAGCCATGTTCGCGCATTAC 1197

Qy        1300 AGGCGTTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCG 1359
               ||    | || ||    ||| | |||||||| |||||  | || || || ||||||  ||
Db        1198 AGAATCTCGCTCAGGAACTGGTCGCGCGCGGTCATCGGGTGACCTTTATTCAGCAATACG 1257

Qy        1360 ATGCCCGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGC 1419
               ||         ||| |      ||||  ||| |  |||  | |||||||   | | |  |
Db        1258 ATATTAAACACTTGATCGATAGCGAAACCATTGGATTTCATTCCGTCGGGACAGACAGCC 1317

Qy        1420 ATCCTGCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGC 1479
               ||||  ||||  || |  ||| |||| | || | |  |||    ||    || | |||  
Db        1318 ATCCCCCCGGCGCGTTAACGCGCGTGCTACACCTGGCGGCTCATCCTCTGGGGCCGTCAA 1377

Qy        1480 TGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTG 1539
               || |     | ||| | ||  | |||  |  |||||||||||||||||||||||| ||  
Db        1378 TGCTGAAGCTCATCAATGAAATGGCGCGCACCACCGATATGCTGTGCCGCGAACTCCCCC 1437

Qy        1540 CGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCG 1599
                || | | || |   ||  | |||||||||| || |  ||  |||||||| |||| ||||
Db        1438 AGGCATTTAACGATCTGGCCGTCGATGGCGTCATTGTTGATCAAATGGAACCGGCAGGCG 1497

Qy        1600 GATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCA 1659
                  | || |||||||| |||   ||||||||| | || || |||||||| ||||  |||
Db        1498 CGCTCGTTGCTGAAGCACTGGGACTGCCGTTTATCTCTGTCGCCTGCGCGCTGCCTCTCA 1557

Qy        1660 ATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAG 1719
               |||||||| | |  || || || ||||| ||||| |||   |  |        |||   |
Db        1558 ATCGTGAACCGGATATGCCCCTGGCGGTTATGCCTTTCGAATACGGGACCAGCGACGCGG 1617

Qy        1720 CGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCG 1779
               | |   |||||| |   |||||     | || |||||  |  | |||||||||||| | | 
Db        1618 CTCGCGAACGTTATGCCGCCAGTGAAAAAATTTATGACTGGCTAATGCGTCGTCATGACC 1677

Qy        1780 ACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGT 1839
                 || ||   | |||||   |      |    ||  |    ||||      | || ||||
Db        1678 GTGTCATTGCCGAACACAGCCACAGAATGGGCTTAGCCCCCCGGCAAAAGCTTCACCAGT 1737

Qy        1840 GCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGC 1899
```

```
Db       1738 GTTTTTCGCCACTGGCGCAAATCAGCCAGCTTGTTCCTGAACTGGATTTTCCCCGCAAAG 1797

Qy       1900 AACTGCCCGCCTGCTATCACGCCGTGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGC 1959

Db       1798 CGTTACCGGCTTGTTTTCATGCCGTCGGGCCTCTGCGCGAAACGCACGCACCGTCAACGT 1857

Qy       1960 TCCATGCGCCCTG---GCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGC 2016

Db       1858 CTTCATCCCGTTATTTTACATCCTCAGAAAAACCCCGGATTTTCGCCTCGCTGGGCACGC 1917

Qy       2017 TGCAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGC 2076

Db       1918 TTCAGGGACACCGTTATGGGCTGTTTAAAACGATAGTGAAAGCCTGTGAAGAAATTGACG 1977

Qy       2077 TATCGCTGGTGATCGCCCATTGTGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGC 2136

Db       1978 GTCAGCTCCTGTTAGCCCACTGTGGTCGTCTTACGGACTCTCAGTGTGAAGAGCTGGCGC 2037

Qy       2137 TCGCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGC 2196

Db       2038 GAAGCCGTCATACACAGGTGGTGGATTTTGCCGATCAGTCAGCCGCGCTGTCTCAGGCGC 2097

Qy       2197 AGCTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGC 2256

Db       2098 AGCTGGCGATCACCCACGGCGGCATGAATACGGTACTGGACGCGATTAATTACCGGACGC 2157

Qy       2257 CGATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGC 2316

Db       2158 CCCTTTTAGCGCTTCCGCTGGCCTTTGATCAGCCCGGCGTCGCGTCACGCATCGTTTATC 2217

Qy       2317 ATGACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGC 2376

Db       2218 ACGGCATCGGCAAGCGTGCTTCCCGCTTTACCACCAGCCATGCTTTGGCTCGTCAGATGC 2277

Qy       2377 AACAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGC 2436

Db       2278 GTTCATTGCTGACCAACGTCGACTTTCAGCAGCGCATGGCGAAAATCCAGACAGCCCTTC 2337

Qy       2437 AGCGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGC 2496

Db       2338 GTTTGGCAGGGGGCACCATGGCCGCTGCCGATATCATTGAGCAGGTTATGTG-CACCGGT 2396

Qy       2497 AAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGC 2556

Db       2397 CAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCGTGGGGGC 2456

Qy       2557 TGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATG 2616

Db       2457 TGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATATGCGTAT 2516

Qy       2617 CCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAG 2676

Db       2517 TTTGCTTATCGACGCCGCACCCCAGGCGGGCGGAATCATACGTGGTCATTTCACCACGA 2576

Qy       2677 CGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGG 2736

Db       2577 TGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACTGGCCCGA 2636

Qy       2737 TTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGC 2796

Db       2637 CTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTTGTATTAC 2696

Qy       2797 ATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACAC 2856

Db       2697 TTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGATGGATAC 2756

Qy       2857 AGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGC 2916

Db       2757 CGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTATCGGTGC 2816

Qy       2917 GCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCA 2976
```

```
Db       2817 CCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGGGCTTCCA 2876

Qy       2977 GGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCT 3036

Db       2877 GGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTCCCATTAT 2936

Qy       3037 GATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAG 3096

Db       2937 CATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGCCGCTCTC 2996

Qy       3097 CGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAA 3156

Db       2997 GCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAGATCCTGA 3056

Qy       3157 CACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCT 3216

Db       3057 ATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTCAGACACT 3116

Qy       3217 GCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTG 3276

Db       3117 GCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACGCATTCTG 3176

Qy       3277 GCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCAC 3336

Db       3177 G---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATCCTACCAC 3233

Qy       3337 CGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCAC 3396

Db       3234 CGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTGATGTCTT 3293

Qy       3397 CGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACA 3456

Db       3294 TACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGCAGCAGCA 3353

Qy       3457 GCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTG 3516

Db       3354 GGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATTCACGCTG 3413

Qy       3517 GCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGG 3576

Db       3414 GCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTTATGCGGG 3473

Qy       3577 GCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGG 3636

Db       3474 AAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTCCGGTATT 3533

Qy       3637 TGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACT 3696

Db       3534 AGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAAACCAACT 3593

Qy       3697 TATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGC 3756

Db       3594 ACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGCTGCGGGG 3653

Qy       3757 ATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAG 3816

Db       3654 ATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGTCTACGAG 3713

Qy       3817 GACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGAA 3876

Db       3714 GATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGCCATTGAA 3773

Qy       3877 GAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACG 3936

Db       3774 GAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCCGGTTACG 3833

Qy       3937 CCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCCG 3996

Db       3834 CCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGATCAAACC 3893

Qy       3997 CTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTT 4056

Db       3894 CGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCGTCAGTTT 3953
```

```
Qy     4057 CTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTT 4116
Db     3954 CTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGTCCCTTTT 4013

Qy     4117 CTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGC 4176
Db     4014 TTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGCATGGAGA 4073

Qy     4177 AGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCC 4236
Db     4074 AGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGCGTTTTCT 4133

Qy     4237 TTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAATT 4296
Db     4134 TTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATACGTTGATA 4193

Qy     4297 CATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTG 4356
Db     4194 CACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGCATTAGTT 4253

Qy     4357 CAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTG 4416
Db     4254 CAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGCCAGAGTC 4313

Qy     4417 AGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGC 4476
Db     4314 AGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGGTCGCAGG 4373

Qy     4477 TTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCTT 4536
Db     4374 TTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGACCTGTTA 4433

Qy     4537 CGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAAC 4596
Db     4434 AGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCATGAGTAAC 4493

Qy     4597 TCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACACC 4656
Db     4494 TCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCATCACACG 4553

Qy     4657 GTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTG 4716
Db     4554 GTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGATGGCCTC 4613

Qy     4717 GCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCG 4776
Db     4614 GCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACTGGCGCCT 4673

Qy     4777 CCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATC 4836
Db     4674 GAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGCGAACCTC 4733

Qy     4837 GACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCAC 4896
Db     4734 GACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGAGCAGCAT 4793

Qy     4897 TACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTTT 4956
Db     4794 TACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTTTGATTTT 4853

Qy     4957 CGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAA 5016
Db     4854 CGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCTTACCCAG 4913

Qy     5017 AGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGT 5076
Db     4914 AGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCTGGTCGGC 4973

Qy     5077 GCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCC 5136
Db     4974 GCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGCGACAGCA 5033
```

```
Qy     5137 AGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACGC 5196
Db     5034 GGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCATGCGGTCG 5090

Qy     5197 AAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGA 5256
Db     5091 AAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGATGCAAAAA 5150

Qy     5257 CGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGATG 5316
Db     5151 CCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTTATTGACG 5210

Qy     5317 GGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGC 5376
Db     5211 ATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAACGTCTGA 5270

Qy     5377 AGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCGT 5436
Db     5271 TGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAACCGGCGT 5330

Qy     5437 TTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGATC 5496
Db     5331 TTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCGTTTGATC 5390

Qy     5497 ATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACGC 5556
Db     5391 ATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGC 5450

Qy     5557 TGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCG 5616
Db     5451 TGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATCATGGGCG 5510

Qy     5617 TGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTA 5676
Db     5511 TGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAGTTGACCA 5570

Qy     5677 ACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCT 5736
Db     5571 ATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCGGCAAGCT 5630

Qy     5737 GGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCGC 5796
Db     5631 GGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGTCAGGCGC 5690

Qy     5797 TCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGAT 5856
Db     5691 TGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCTGCCACAG 5750

Qy     5857 CCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTT 5916
Db     5751 CCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAGCAGGTTT 5810

Qy     5917 ATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGC 5976
Db     5811 ACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAGCGGCAGT 5870

Qy     5977 GCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTT 6036
Db     5871 CAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCCCTTACTT 5930

Qy     6037 CGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTT 6096
Db     5931 CCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC------TCT 5984

Qy     6097 ACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAA 6156
Db     5985 AGCGCCATGTCTTTCCCGGAGCGTCGCCTGAAGTTTTGACAGGGGCGGCGCATAGAGGAA 6044

Qy     6157 ACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGTA 6216
Db     6045 GCCAAAAGAAACACAACCTTCTTTGCCCCTGACGGCGTGATGCATACGGTGCGCCATATA 6104

Qy     6217 TAAGCGCTTAAGATAGCCTTTGCGCGGATATAGCGGAACGGCCAGCGTTGATGCACCAG 6276
```

```
                  || || || || ||||| ||||| || ||||||||||| ||||| ||||||||||||||
Db          6105 CAACCGTTTGAGGTAGCCCTTGCGTGGAATATAGCGGAATGGCCAACGTTGATGCACCAG 6164

Qy          6277 GCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACTG 6336
                  || ||||||||||| || |||||    || |||| |||||| || || |||||||||||
Db          6165 CCCGTCGTGCACCATAAAATAGAGTAATCCATACGCCGTCATACCTGCGCCAATCCACTG 6224

Qy          6337 CAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACAC 6396
                  ||||||||||| |||      || | |||||||||||  ||||  | | | ||||| ||
Db          6225 GAGCGGCCACATTCCTGTACTGCCCAGATAAATCAGCAGGATCGATAATGCAGCAAAAAC 6284

Qy          6397 CACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGATG 6456
                 ||| |||||||||||||| |  ||||||  |||  |  |||||||||| || || |||||
Db          6285 CACGGCATAAAGATCGTTAACTTCAAACGCACCTTTACGCGGTTCATGATGTGAAAGATG 6344

Qy          6457 CCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCAT 6516
                 ||| ||||| ||||| ||||||||||||||||| || | ||| || || |  | ||||||
Db          6345 CCATCCCCAACCCCAGCCGTGCATGATGTATTTGTGTGCCAGTGCAGCAATCACTTCCAT 6404

Qy          6517 CACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATTGTTCGTCCATT 6576
                   | | ||||||  | || | ||| ||  | ||||| | ||| ||| ||  |||  |
Db          6405 GCCAATCACGGTAACGAAAACGATCAGGGCATTCCAAATCCACAACATAATTTCTCCGGT 6464

Qy          6577 TGTGGAAAAGGGAAGTA 6593
                  | |      || || |
Db          6465 AGAGACGTCTGGCAGCA 6481


RESULT 4
AY166713
LOCUS       AY166713                6586 bp    DNA     linear   BCT 18-DEC-2002
DEFINITION  Pantoea stewartii zeaxanthin-diglucoside biosynthetic operon,
            complete sequence.
ACCESSION   AY166713
VERSION     AY166713.1  GI:27228289
KEYWORDS    .
SOURCE      Pantoea stewartii
  ORGANISM  Pantoea stewartii
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1  (bases 1 to 6586)
  AUTHORS   deSouza,M.L., Kollmann,S.R. and Schroeder,W.A.
  TITLE     Carotenoid Biosynthesis (WO 02/079395 A2)
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6586)
  AUTHORS   deSouza,M.L., Kollmann,S.R. and Schroeder,W.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2002) Biotechnology Development Center, Cargill
            Inc., P.O. Box 5702, Minneapolis, MN 55440, USA
FEATURES             Location/Qualifiers
     source          1. .6586
                     /organism="Pantoea stewartii"
                     /mol_type="genomic DNA"
                     /strain="ATCC 8200"
                     /db_xref="ATCC:8200"
                     /db_xref="taxon:66269"
     gene            169. .6399
                     /gene="zeaxanthin-diglucoside biosynthetic operon"
     gene            169. .1080
                     /gene="crtE"
     CDS             169. .1080
                     /gene="crtE"
                     /codon_start=1
                     /transl_table=11
                     /product="geranylgeranyl pyrophosphate synthase"
                     /protein_id="AAN85596.1"
                     /db_xref="GI:27228290"
                     /translation="MMVCAKKHVHLTGISAEQLLADIDSRLDQLLPVQGERDCVGAAM
                     REGTLAPGKRIRPMLLLLTARDLGCAISHGGLLDLACAVEMVHAASLILDDMPCMDDA
                     QMRRGRPTIHTQYGEHVAILAAVALLSKAFGVIAEAEGLTPIAKTRAVSELSTAIGMQ
                     GLVQGQFKDLSEGDKPRSADAILLTNQFKTSTLFCASTQMASIAANASCEARENLHRF
                     SLDLGQAFQLLDDLTDGMTDTGKDINQDAGKSTLVNLLGSGAVEERLRQHLRLASEHL
                     SAACQNGHSTTQLFIQAWFDKKLAAVS"
     gene            1090. .2385
                     /gene="crtX"
     CDS             1090. .2385
                     /gene="crtX"
                     /codon_start=1
                     /transl_table=11
                     /product="zeaxanthin glucosyl transferase"
                     /protein_id="AAN85597.1"
                     /db_xref="GI:27228291"
                     /translation="MSHFAVIAPPFFSHVRALQNLAQELVARGHRVTFFQQHDCKALV
                     TGSDIGFQTVGLQTHPPGSLSHLLHLAAHPLGPSMLRLINEMARTSDMLCRELPAAFH
                     ALQIEGVIVDQMEPAGAVVAEASGLPFVSVACALPLNREPGLPLAVMPFEYGTSDAAR
                     ERYTTSEKIYDWLMRRHDRVIAHHACRMGLAPREKLHHCFSPLAQISQLIPELDFPRK
                     ALPDCFHAVGPLRQPQGTPGSSTSYFPSPDKPRIFASLGTLQGHRYGLFRTIAKACEE
                     VDAQLLLAHCGGLSATQAGELARGGDIQVVDFADQSAALSQAQLTITHGGMNTVLDAI
                     ASRTPLLALPLAFDQPGVASRIVYHGIGKRASRFTTSHALARQIRSLLTNTDYPQRMT
                     KIQAALRLAGGTPAAADIVEQAMRTCQPVLSGQDYATAL"
     gene            2369. .3517
                     /gene="crtY"
     CDS             2369. .3517
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /product="lycopene beta-cyclase"
                     /protein_id="AAN85598.1"
                     /db_xref="GI:27228292"
                     /translation="MQPHYDLILVGAGLANGLIALRLQQQHPDMRILLIEAGPEAGGN
                     HTWSFHEEDLTLNQHRWIAPLVVHHWPDYQVRFPQRRRHVNSGYYCVTSRHFAGILRQ
                     QFGQHLWLHTAVSAVHAESVQLADGRIIHASTVIDGRGYTPDSALRVGFQAFIGQEWQ
                     LSAPHGLSSPIIMDATVDQQNGYRFVYTLPLSATALLIEDTHYIDKANLQAERARQNI
                     RDYAARQGWPLQTLLREEQGALPITLTGDNRQFWQQQPQACSGLRAGLFHPTTGYSLP
                     LAVALADRLSALDVFTSSSVHQTIAHFAQQRWQQQGFFRMLNRMLFLAGPAESRWRVM
                     QRFYGLPEDLIARFYAGKLTVTDRLRILSGKPPVPVFAALQAIMTTHR"
     gene            3529. .5007
                     /gene="crtI"
     CDS             3529. .5007
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene desaturase"
                     /protein_id="AAN85599.1"
                     /db_xref="GI:27228293"
                     /translation="MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYQ
                     EQGFTFDAGPTVITDPSAIEELFALAGKQLKDYVELLPVTPFYRLCWESGKVFNYDND
                     QAQLEAQIQQFNPRDVAGYRAFLDYSRAVFNEGYLKLGTVPFLSFKDMLRAAPQLAKL
                     QAWRSVYSKVAGYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLIHALEREWGVWFPRG
                     GTGALVNGMIKLFQDLGGEVVLNARVSHMETVGDKIQAVQLEDGRRFETCAVASNADV
                     VHTYRDLLSQHPAAAKQAKKLQSKRMSNSLFVLYFGLNHHHDQLAHHTVCFGPRYREL
                     IHEIFNHDGLAEDFSLYLHAPCVTDPSLAPEGCGSYYVLAPVPHLGTANLDWAVEGPR
                     LRDRIFDYLEQHYMPGLRSQLVTHRMFTPFDFRDELNAWQGSAFSVEPILTQSAWFRP
                     HNRDKHIDNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI"
     gene            5043. .5933
                     /gene="crtB"
     CDS             5043. .5933
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene synthase"
                     /protein_id="AAN85600.1"
                     /db_xref="GI:27228294"
                     /translation="MAVGSKSFATASTLFDAKTRRSVLMLYAWCRHCDDVIDDQTLGF
                     HADQPSSQMPEQRLQQLEMKTRQAYAGSQMHEPAFAAFQEVAMAHDIAPAYAFDHLEG
                     FAMDVRETRYLTLDDTLRYCYHVAGVVGLMMAQIMGVRDNATLDRACDLGLAFQLTNI
                     ARDIVDDAQVGRCYLPESWLEEEGLTKANYAAPENRQALSRIAGRLVREAEPYYVSSM
                     AGLAQLPLRSAWAIATAKQVYRKIGVKVEQAGKQAWDHRQSTSTAEKLTLLLTASGQA
                     VTSRMKTYPPRPAHLWQRPI"
     gene            complement(5872. .6399)
                     /gene="crtZ"
     CDS             complement(5872. .6399)
                     /gene="crtZ"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-carotene hydroxylase"
                     /protein_id="AAN85601.1"
                     /db_xref="GI:27228295"
                     /translation="MLWIWNALIVFVTVVGMEVVAALAHKYIMHGWGWGWHLSHHEPR
                     KGAFEVNDLYAVVFAIVSIALIYFGSTGIWPLQWIGAGMTAYGLLYFMVHDGLVHQRW
                     PFRYIPRKGYLKRLYMAHRMHHAVRGKEGCVSFGFLYAPPLSKLQATLRERHAARSGA
                     ARDEQDGVDTSSSGK"
ORIGIN

  Query Match             37.4%;  Score 2620;  DB 1;  Length 6586;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4016;  Conservative    0;  Mismatches 2170;  Indels   24;  Gaps    7;

Qy        373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
Db        220 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 279

Qy        433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
Db        280 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 339

Qy        493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
Db        340 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 399

Qy        553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
Db        400 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 459

Qy        613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
Db        460 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 519

Qy        673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
Db        520 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 579

Qy        733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
Db        580 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 639

Qy        793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
Db        640 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 699

Qy        853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
Db        700 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 759

Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
Db        760 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 819

Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
Db        820 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 879

Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
Db        880 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 939

Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
Db        940 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 999

Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
Db       1000 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 1059

Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
Db       1060 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 1114

Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
Db       1115 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 1174

Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
Db       1175 GTCATCGTGTTACGTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 1234

Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
Db       1235 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 1294

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
Db       1295 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 1354

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
Db       1355 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 1414

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
Db       1415 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 1474

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
Db       1475 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 1534

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
Db       1535 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 1594

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
Db       1595 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 1654

Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
Db       1655 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 1714

Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
Db       1715 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 1774

Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
Db       1775 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTCCGTCCCCGGACA 1834

Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
Db       1835 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 1894

Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
Db       1895 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 1954

Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
Db       1955 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 2014

Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
Db       2015 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGATGAATA 2074

Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
Db       2075 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 2134

Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
Db       2135 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 2194

Qy       2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
```

```
Db        2195 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 2254

Qy        2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466

Db        2255 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 2314

Qy        2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526

Db        2315 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 2373

Qy        2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586

Db        2374 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 2433

Qy        2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646

Db        2434 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 2493

Qy        2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706

Db        2494 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 2553

Qy        2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766

Db        2554 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 2613

Qy        2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826

Db        2614 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 2673

Qy        2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886

Db        2674 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 2733

Qy        2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946

Db        2734 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 2793

Qy        2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006

Db        2794 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 2853

Qy        3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066

Db        2854 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 2913

Qy        3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126

Db        2914 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 2973

Qy        3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186

Db        2974 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 3033

Qy        3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246

Db        3034 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 3093

Qy        3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306

Db        3094 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 3150

Qy        3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366

Db        3151 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 3210

Qy        3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426

Db        3211 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 3270

Qy        3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486

Db        3271 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 3330
```

```
Qy        3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546

Db        3331 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 3390

Qy        3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606

Db        3391 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 3450

Qy        3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666

Db        3451 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 3510

Qy        3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726

Db        3511 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 3570

Qy        3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786

Db        3571 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 3630

Qy        3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846

Db        3631 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTGATGCAGGCCCT 3690

Qy        3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906

Db        3691 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 3750

Qy        3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966

Db        3751 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 3810

Qy        3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026

Db        3811 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 3870

Qy        4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086

Db        3871 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 3930

Qy        4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146

Db        3931 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 3990

Qy        4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206

Db        3991 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 4050

Qy        4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266

Db        4051 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGAATCCG 4110

Qy        4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326

Db        4111 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 4170

Qy        4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386

Db        4171 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 4230

Qy        4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446

Db        4231 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 4290

Qy        4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506

Db        4291 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 4350

Qy        4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566

Db        4351 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 4410
```

```
Qy          4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
                 | |   ||| |    ||||| ||||| ||||| |||||||||||||||||||| || ||
Db          4411 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 4470

Qy          4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
                 || |  || || || ||||| || || |||||||||||||| || || || |||||  ||
Db          4471 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 4530

Qy          4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
                 ||  | || ||||| |||         ||||| || |||||||| |||||  | |||||
Db          4531 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 4590

Qy          4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
                 || ||   ||  |||||||| |||||||||   || |||||||||| ||||||| | |||
Db          4591 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 4650

Qy          4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
                 || || || ||  | ||||| ||  || ||||||||      |||||||| ||  |||||
Db          4651 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 4710

Qy          4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
                 ||||| |||||||  || || ||||| || |||||||| || || ||    || || |||
Db          4711 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 4770

Qy          4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
                 || ||| | |||||||||||||| ||||| ||||||  |||  |||||   || || |||
Db          4771 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 4830

Qy          4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
                 || || ||| | || || ||| |||| || |||||||||||||| || ||||||||||||
Db          4831 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 4890

Qy          5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
                     | ||    ||||| |||||||| || || || ||||| ||||||||| | |||
Db          4891 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 4950

Qy          5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
                 ||||| ||||| ||||| ||||| || ||  |  | ||||||||||| |       | ||
Db          4951 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 5010

Qy          5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
                 ||     | | |||||  | || |      ||||||||||||| |||||||| || || |
Db          5011 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 5067

Qy          5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
                 | || ||  | | ||| || || ||    || || ||||||  ||||||||| || || |
Db          5068 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 5127

Qy          5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
                 |||| || |||||||| || || ||||| |  ||||| ||||||      | |   ||||
Db          5128 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 5187

Qy          5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
                    |     |     |  ||   ||  ||||||| ||  ||||  |||| || |  ||||
Db          5188 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 5247

Qy          5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
                 ||  |||  |||| ||| | || || || |||   ||||||||||| || |||||    |
Db          5248 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 5307

Qy          5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
                 |  |  |  | |        ||||| || ||||||||||| || || |||||||  |||
Db          5308 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 5367

Qy          5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
                 |    |  |||  ||   | || || ||||||||||| |||||||||||||| || || |
Db          5368 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 5427

Qy          5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
                 | ||  ||||||||||||   || ||||||| |||||  | ||   |||||||| |||||
Db          5428 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 5487

Qy          5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
                 ||||| | || ||||| |||||| | || ||||||||||| || ||||| || |||||
Db          5488 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 5547

Qy          5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
                 |    || ||||| ||||||||  ||  |||||| ||  |||  ||| |       |  |
Db          5548 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 5607

Qy          5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                       |||| ||  || | ||    ||  |   |    | || | |||  | ||   ||
Db          5608 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 5667

Qy          5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
                 | ||||||||||||||    ||  |     ||||| |  |    |||||  | |||||||
Db          5668 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 5727

Qy          5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
                 | ||||||||||| || ||       || || ||  |||||||||| |||||| | || |
Db          5728 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 5787

Qy          5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
                 |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db          5788 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 5847

Qy          6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
                 || |||  |   | |||  |||  | ||||| ||  ||        |  ||| ||||| |
Db          5848 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 5907

Qy          6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
                 ||  ||| |||||||| ||        | ||    | || |  |  |  | |||||| || ||
Db          5908 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 5961

Qy          6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
                  || |||   | |||||||||||| || |||||||| ||||||||||| ||    || |
Db          5962 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 6021

Qy          6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
                  || ||||||||||| ||||| ||||||||||| || || || |||||||||||||| ||
Db          6022 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 6081

Qy          6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
                  ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db          6082 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 6141

Qy          6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
                  || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db          6142 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 6201

Qy          6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
                 ||||||  |||||| ||  |  || || ||||| ||||| |||||||| |  |||||
Db          6202 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 6261

Qy          6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
                  ||  |  |||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db          6262 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 6321

Qy          6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
                 |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db          6322 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 6381

Qy          6547 GTTCCATAACCAGAGCATTGTTCGTCCATT 6576
                  ||||| | ||| | |||  |   |||| |
Db          6382 ATTCCAAATCCACAACATAATCTCTCCAGT 6411


RESULT 5
ERWCRTA
LOCUS       ERWCRTA                 6965 bp    DNA     linear   BCT 11-APR-2001
DEFINITION  Pantoea agglomerans CrtE (crtE), CrtX (crtX), dycopene cyclase
            (crtY), phytoene dehydrogenase (crtI), and prephytoene
            pyrophosphate synthase (crtB) genes, complete cds.
ACCESSION   M90698
VERSION     M90698.1  GI:148393
KEYWORDS    .
```

```
SOURCE      Pantoea agglomerans
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1  (bases 1101 to 1300; 4301 to 4410; 4641 to 4750; 4881 to 5000;
            5031 to 5140; 5561 to 5660; 6031 to 6160)
  AUTHORS   Liu,S.T.
  TITLE     Carotenoid-biosynthesis genes as a genetic marker for the purpose
            of gene cloning
  JOURNAL   Biochem. Biophys. Res. Commun. 195 (1), 259-263 (1993)
  MEDLINE   93371414
   PUBMED   8395826
REFERENCE   2  (bases 1 to 6965)
  AUTHORS   To,K.Y., Lai,E.M., Lee,L.Y., Lin,T.P., Hung,C.H., Chen,C.L.,
            Chang,Y.S. and Liu,S.T.
  TITLE     Analysis of the gene cluster encoding carotenoid biosynthesis in
            Erwinia herbicola Eho13
  JOURNAL   Microbiology 140 (Pt 2), 331-339 (1994)
  MEDLINE   94236237
   PUBMED   8180698
FEATURES             Location/Qualifiers
     source          1. .6965
                     /organism="Pantoea agglomerans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:549"
     mRNA            776. .6687
                     /product="CrtE"
     gene            926. .1834
                     /gene="crtE"
     CDS             926. .1834
                     /gene="crtE"
                     /codon_start=1
                     /transl_table=11
                     /product="CrtE"
                     /protein_id="AAA21260.1"
                     /db_xref="GI:148394"
                     /translation="MTVCAKKHVHPTRSAAEQLLADIDRRLEQLLPVEGERDFVGAAM
                     REGALAPGKRIRPMLLLLTARDLGCAVSHEGLLDIACAVEMVHAASLILDDMPCMDDA
                     QIVGGRPTVHCQYGEHVAILAAVALLSKAFGVIADADGLTALAKNRAVSELSNAIGMQ
                     GLVQGQFKDLSEGDKRRSAEAILMTNHFKTSTLFCASMQMASIVANASSERRDYLHRF
                     SLDLGQAFQLLDDLTDGMADTGKDINQDEGKSTLVNLLGSRAVEERLRHSLHLASEHL
                     SAACQNGHSTQQFIQAWFDKKLAAVS"
     gene            1844. .3139
                     /gene="crtX"
     CDS             1844. .3139
                     /gene="crtX"
                     /codon_start=1
                     /transl_table=11
                     /product="CrtX"
                     /protein_id="AAA21261.1"
                     /db_xref="GI:148395"
                     /translation="MSHFAAIAPPFYSHVRALENLAQELVARGHQVTFIQQQDTKPLI
                     QHPHIGFHAVGEESHPSGSLSRLLHLAAHPLGPSMLRLINEMARTSDMLCLALPRAFN
                     DLAIDGVIVDQMEPAGALVAEALGLPFVSVACALPLNREPDLPLAVMPFDYGTSDAAR
                     ERYAASEKIYDWLMRRHDRVIAHHSHRMGLAPRDKLHQCFSPLAQISQLIPELDFPRK
                     ALPACFHAVGPLRETQTTSTSSPLYFSPSEKPRIFASLGTLQGHRYGLFKAIVKACEE
                     IDGQLLLAHCGRLTAFQAEELARSRHTQVVDFADQSAALSQAQLAITHGGMNTVLDAI
                     NYRTPLLALPLAFDQPGVASRIVSHGIGKRASRFTTSHALARHMRALLTNVDYPQRMT
                     KIQTALRLAGGTMAAADIVEQAMRTGQPVLKGSGYATAL"
     gene            3123. .4271
                     /gene="crtY"
     CDS             3123. .4271
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /product="dycopene cyclase"
                     /protein_id="AAA21262.1"
                     /db_xref="GI:148396"
                     /translation="MRPHYDLILVGAGLANGLIALRLQQQQPDMRILLIEAAPQAGGN
                     HTWSFHDADLTESQHRWVAPLVVYHWPDYQVRFPTRRRKLNSGYFSVTSQRFAEVLQQ
                     KFGQHLWISRAVAEVHADAVRLNNGQVISASAVIDGRGYTPNSALNVGFQAFIGQEWR
                     LSKPHGLSSPIIMDATVDQQNGYRFVYSLPLSATELLIEDTHYIDNATLEPERARQNI
                     RDYAAQQDWQLQTLLREEQGALPITLTGDSTAFWQQQPLACSGLRAGLFHPTTGYSLP
                     LAVALADRLSALDVFTSSSIHQAITHFAHERWQQQRFFRMLNRMLFLAGPADSRWRVM
                     QRFYGLPEDLISRFYAGKLTLTDRLRILSGKPPVPVLAALQAIMTTHR"
     gene            4283. .5761
                     /gene="crtI"
     CDS             4283. .5761
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene dehydrogenase"
                     /protein_id="AAA21263.1"
                     /db_xref="GI:148397"
                     /translation="MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYE
                     DQGFTFDAGPTVITDPSAIEELFTLAGKQLKDYVELLPVAPFYRLCWESGKVFNYDND
                     QAQLEAQIQQFNPRDVEGYRQFLDYSRAVFKEGYLKLGTVPFLSFRDMLRAAPQLAKL
                     QAWRTVYSKVASYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLIHALEREWGVWFPRG
                     GTGALVKGMIKLFQDLGGEVVLNAKVSHMETTGDTIEAVHLEDGRRFPTRAVASNADV
                     VHTYRDLLSQHPAAVKQSKKLQTKRMSNSLFVLYFGLNHHHDQLAHHTVCFGPRYREL
                     IHEIFNHDGLADDFSLYLHAPCVTDSSLAPEGCGSYYVLAPVPHLGTANLDWTVEGPR
                     LRDRIFEYLEQHYMPGLRSQLVTQRMFTPFDFRDQLNAYQGSAFSVEPVLTQSAWFRP
                     HNRDKTINNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI"
     gene            5797. .6687
                     /gene="crtB"
     CDS             5797. .6687
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /product="prephytoene pyrophosphate synthase"
                     /protein_id="AAA21264.1"
                     /db_xref="GI:148398"
                     /translation="MAVGSKSFATASKLFDAKTRRSVLMLYSWCRHCDDVIDDQTLGF
                     QTDQPSLQTPEQRLMQLEMKTRQAYAGSQMHEPAFAAFQEVAMAHDIAPAYAFDHLEG
                     FAMDVREAEYIQLDDTLRYCYHVAGVVGLMMAQIMGVRDNATLDRACDLGLAFQLTNI
                     ARDIVEDAHVGRCYLPASWLENEGLNKENYAAPENRQALSRIARRLVQEAEPYYLCAT
                     AGLAGLPLRSAWAIATAKQVYRKIGVKVEQAGTEAWDHRQSTTTPEKLSLLLMASGQA
                     ITSRMRPHPPRPAHLWQRPI"
ORIGIN

  Query Match             35.3%;  Score 2468.8;  DB 1;  Length 6965;
  Best Local Similarity   64.0%;  Pred. No. 0;
  Matches 3829;  Conservative    0;  Mismatches 2132;  Indels   21;  Gaps    6;

Qy          396 GCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCG 455
                ||  ||||| || ||| |||||| ||      || | |||   ||  |||||||||||||
Db         1000 GCGGCTTGAGCAGTTATTGCCTGTCGAGGGAGAACGGGATTTTGTAGGTGCCGCGATGCG 1059

Qy          456 TGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCG 515
                ||  ||  ||||||| | |||  ||||||||||| ||  |  |  || ||||  |||| ||
Db         1060 TGAGGGCGCGCTGGCACCGGGTAAACGTATTCGCCCGATGCTGTTGTTGCTCACAGCACG 1119

Qy          516 CGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAAT 575
                |||| |||| ||||   | |  ||  | ||  | || ||| ||||||| || ||||||||
Db         1120 CGATCTGGGCTGCGCTGTCAGCCACGAGGGATTGCTGGATATCGCCTGCGCCGTGGAAAT 1179

Qy          576 GGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATGGATAACGCGCAGATGCG 635
                |||||||||||||||| ||||||| ||||| || ||||  |||||| | ||||||||
Db         1180 GGTGCACGCGGCATCGTTGATTCTTGATGATATGCCCTGTATGGATGATGCGCAGATCGT 1239

Qy          636 TCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCGCCAT 695
                  | || ||||||||||| ||  |  | |  || ||  | ||||| || || || ||  |
Db         1240 CGGCGGCCGCCCTACCGTTCACTGTCAGTACGGCGAGCATGTGGCAATACTGGCGGCGGT 1299

Qy          696 CGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCATACA 755
                |||  | || |||   || ||||  || |||||   ||||   ||  || | ||  |
Db         1300 CGCCTTACTGAGCAAAGCCTTTGGGGTCATTGCTGATGCAGATGGCCTGACCGCGCTGGC 1359

Qy          756 TAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGG 815
                 |||      ||  |  | ||| | ||    ||| | ||  |||| || || || || ||
Db         1360 AAAAAACCGCGCAGTGTCAGAATTATCAAACGCCATTGGTATGCAAGGTTTGGTACAGGG 1419

Qy          816 GCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGAC 875
                 || ||  ||||||||   || |||    || |||||||| | ||||| ||    |||||
Db         1420 TCAGTTTAAGGATCTGTCAGAAGGCGATAAGCGCCGCAGCGCTGAAGCCATTTTGATGAC 1479
```

```
Qy        876 CAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGC 935
Db       1480 GAACCACTTTAAAACCAGCACCCTGTTTTGCGCGTCTATGCAGATGGCCTCAATTGTGGC 1539

Qy        936 TGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGC 995
Db       1540 GAATGCCTCCAGCGAACGACGCGATTACCTACACCGTTTTTCGCTTGATCTCGGCCAAGC 1599

Qy        996 GTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGCACCA 1055
Db       1600 GTTTCAACTGCTGGACGACTTAACGGATGGCATGGCTGACACGGGTAAGGACATCAATCA 1659

Qy       1056 GGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCT 1115
Db       1660 GGATGAAGGCAAATCGACCCTGGTCAATTTATTAGGCTCCAGAGCAGTTGAAGAGCGTCT 1719

Qy       1116 GCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCCTGCGCCTGCCATCGCGGCATCGC 1175
Db       1720 GCGGCACAGTCTGCATCTCGCCAGTGAGCACCTCTCAGCCGCGTGTCAAAACGGTCATTC 1779

Qy       1176 CACTCGCCAATATATGCACGCGCTGTTTAATCAACAGCTAGCGATATTCAACTGAGCGCG 1235
Db       1780 TACTCAACAATTTATCCAGGCCTGGTTTGACAAAAAACTCGCTGCCGTCAGTTAAGGATG 1839

Qy       1236 GCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCG 1295
Db       1840 CTGCA-----TGAGCCACTTCGCGGCGATCGCTCCGCCTTTTACAGCCATGTCCGTGCA 1894

Qy       1296 TTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAA 1355
Db       1895 CTGGAAAACCTCGCTCAGGAGTTAGTGGCTCGCGGTCATCAGGTGACCTTTATTCAGCAA 1954

Qy       1356 GCCGATGCCCGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAG 1415
Db       1955 CAGGACACTAAACCACTCATTCAGCATCCCCACATCGGATTTCATGCTGTAGGAGAAGAA 2014

Qy       1416 ACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTG 1475
Db       2015 AGCCATCCGTCTGGCTCATTATCGCGCCTGCTGCACTTAGCGGCCCATCCTCTGGGGCCG 2074

Qy       1476 TCGCTGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTG 1535
Db       2075 TCAATGCTTCGGCTCATCAATGAAATGGCGCGCACCAGCGATATGCTGTGCCTGGCGCTC 2134

Qy       1536 CCTGCGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCG 1595
Db       2135 CCCAGAGCCTTTAACGATTTGGCGATAGATGGTGTCATCGTTGATCAGATGGAGCCAGCA 2194

Qy       1596 GGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCG 1655
Db       2195 GGCGCTCTGGTCGCTGAAGCGCTGGGGCTGCCGTTTGTCTCGGTTGCCTGTGCACTCCCG 2254

Qy       1656 GTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGAC 1715
Db       2255 CTTAATCGCGAACCGGATTTGCCCCTGGCGGTGATGCCTTTTGACTATGGCACCAGCGAT 2314

Qy       1716 AAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCAC 1775
Db       2315 GCCGCACGTGAGCGCTATGCAGCCAGCGAAAAAATTTATGACTGGCTGATGCGTCGGCAT 2374

Qy       1776 GGCGACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACAT 1835
Db       2375 GACCGCGTCATTGCCCATCACAGTCACAGGATGGGCTTAGCACCCCGGGATAAACTCCAT 2434

Qy       1836 CAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGT 1895
Db       2435 CAGTGTTTTTCGCCATTGGCACAAATCAGTCAGCTTATTCCAGAACTGGATTTTCCCCGT 2494

Qy       1896 CAGCAACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCG 1955
Db       2495 AAAGCGCTACCGGCTTGTTTTCATGCTGTTGGCCCCCTGCGAGAAACGCAGACGACATCA 2554

Qy       1956 CCGCTCCATGCGCCCTGGCCAGCGCTGCGT---CAGCCGGTGGTTTATGCCTCGCTGGGT 2012
Db       2555 ACCTCTTCGCCGCTTTATTTTTCGCCGTCTGAAAAGCCCCGAATTTTCGCCTCGCTGGGC 2614

Qy       2013 ACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTG 2072
Db       2615 ACGTTGCAGGGTCATCGCTACGGTCTGTTTAAAGCCATCGTGAAAGCCTGTGAAGAGATT 2674

Qy       2073 CGGCTATCGCTGGTGATCGCCCATTGTGGGGATTAAACGCCGAACAGACGCATCAGCTG 2132
Db       2675 GATGGTCAGCTTCTGTTGGCGCACTGTGGCCGCCTGACTGCCTTTCAGGCCGAAGAGCTG 2734

Qy       2133 GAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCAC 2192
Db       2735 GCGCGAAGCCGTCACACGCAGGTTGTGGATTTTGCCGATCAGTCGGCCGCGCTGTCGCAG 2794

Qy       2193 GCGCAGCTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGT 2252
Db       2795 GCGCAGCTTGCCATTACCCATGGCGGGATGAATACGGTGCTGGATGCCATTAATTATCGC 2854

Qy       2253 ACGCCGATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAG 2312
Db       2855 ACGCCCCTTTTAGCCCTTCCGTTGGCATTTGATCAGCCCGGCGTGGCCTCACGTATTGTT 2914

Qy       2313 TGGCATGACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCAT 2372
Db       2915 TCTCATGGGATAGGAAAACGTGCGTCCCGTTTTACCACCAGCCATGCGCTGGCGCGCCAC 2974

Qy       2373 CTGCAACAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAG 2432
Db       2975 ATGCGTGCATTGCTGACCAACGTCGACTACCCGCAACGCATGACGAAAATCCAGACCGCC 3034

Qy       2433 CTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAG 2492
Db       3035 CTTCGTTTAGCCGGCGGCACAATGGCCGCTGCAGATATCGTTGAGCAGGCCATGCG-CAC 3093

Qy       2493 CAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCG 2552
Db       3094 CGGCCAGCCCGTTTTGAAGGGGAGCGGCTATGCGACCGCATTATGATCTGATTTTAGTGG 3153

Qy       2553 GTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGA 2612
Db       3154 GCGCCGGGCTGGCTAATGGCCTCATTGCCCTGCGCCTTCAGCAGCAGCAACCTGACATGC 3213

Qy       2613 AATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATC 2672
Db       3214 GTATTCTGCTTATCGAAGCGGCCCCCCAGGCGGGCGGTAATCATACCTGGTCCTTTCACG 3273

Qy       2673 ACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGT 2732
Db       3274 ACGCTGATTTAACCGAGAGCCAACATCGCTGGGTAGCACCGCTGGTGGTATATCACTGGC 3333

Qy       2733 CAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCA 2792
Db       3334 CCGACTATCAGGTCCGCTTTCCCACGCGGCGTCGTAAGCTGAACAGCGGCTATTTTTCTG 3393

Qy       2793 TCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAA 2852
Db       3394 TGACTTCGCAGCGTTTTGCCGAGGTTTTACAACAGAAGTTTGGTCAGCACTTATGGATTA 3453

Qy       2853 ACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTG 2912
Db       3454 GCCGCGCGGTGGCAGAGGTTCACGCCGACGCTGTTCGACTGAATAATGGTCAGGTCATCA 3513

Qy       2913 CTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTT 2972
Db       3514 GTGCCAGCGCGGTGATTGATGGGCGAGGTTATACGCCAAATTCAGCGCTGAACGTGGGAT 3573

Qy       2973 ATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGA 3032
Db       3574 TCCAGGCGTTTATCGGACAGGAATGGCGACTCAGTAAACCGCACGGTTTATCTTCACCCA 3633

Qy       3033 TCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGC 3092
```

```
Db       3634 TTATTATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTTGTTTACAGCCTGCCGC 3693

Qy       3093 TCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGG 3152
              |    ||       || || || |||||||||||| ||  |  |  |  |   | ||||
Db       3694 TATCTGCGACAGAACTGTTAATCGAAGATACCCACTATATCGATAACGCGACACTGGAAC 3753

Qy       3153 AGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTA 3212
                 |   ||| || ||  | ||   ||| |||||    ||||| |  |||  |||    |
Db       3754 CTGAACGCGCGCGACAAAATATTCGCGATTATGCTGCCCAGCAGGATTGGCAGCTTCAGA 3813

Qy       3213 CGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGAT 3272
              | ||||| ||||| || || |||  ||| || || ||| | |  ||| | | |      |
Db       3814 CTCTGCTTCGTGAGGAACAGGGCGCATTGCCAATAACCTTAACGGGCGATAGCACCGCGT 3873

Qy       3273 TCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCA 3332
              | ||||| || ||||        ||| |||||| || ||||| ||||| ||||||||  |||
Db       3874 TTTGGCAGCAACAGC---CTTTAGCCTGCAGTGGACTGCGTGCCGGACTGTTTCACCCCA 3930

Qy       3333 CCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGC 3392
              | |||||||| ||  | ||    || || ||||||| || ||     |    |||||
Db       3931 CAACCGGTTATTCACTCCCCCTTGCGGTTGCGCTGGCCGATCGCCTGAGTGCGCTTGATG 3990

Qy       3393 CCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCG 3452
               |       |   | |  |    ||    ||    | ||||||||   |||  ||||
Db       3991 TCTTTACATCGTCCTCAATTCATCAGGCGATTACCCACTTTGCCCACGAGCGCTGGCAGC 4050

Qy       3453 AACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGC 3512
              | || || ||||||||  || | || |||||||||||||||| || ||     |  |    |
Db       4051 AGCAACGCTTTTTCCGCATGTTGAATCGCATGCTGTTTTTAGCGGGACCCGCCGATTCAC 4110

Qy       3513 GCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACG 3572
              ||||||| || ||||| || || ||  |  |   ||  |  |||||   ||| ||||| |
Db       4111 GCTGGCGTGTTATGCAGCGCTTCTATGGTTTACCTGAAGATTTAATCTCCCGTTTTTATG 4170

Qy       3573 CCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCA 3632
              | ||  ||||    |||  ||||      || |||||| |||||||||||||| || ||
Db       4171 CGGGCAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTCCGG 4230

Qy       3633 TCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAAACG 3692
              |    |  ||  |||  ||  |  |||   ||        | |   || || ||||||||
Db       4231 TATTAGCGGCATTGCAGGCAATTATGACGACTCATCGTTAAAGAGCGACAACATGAAACC 4290

Qy       3693 CACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGC 3752
               |||   || ||||| ||||||||||| ||||||||  | || ||||| ||||| |||||
Db       4291 AACTACGGTAATTGGTGCAGGCTTTGGTGGCCTGGCATTAGCAATTCGTCTGCAGGCGGC 4350

Qy       3753 GGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTT 3812
              ||| || ||   |    | || ||||| ||||||||||| || || || || ||||| |
Db       4351 GGGGATCCCTGTCTTACTGCTTGAGCAACGCGACAAACCCGGTGGCCGGGCTTATGTCTA 4410

Qy       3813 TGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCAT 3872
              ||| ||    || || || || ||||| || || || || || ||||||||||||||||||
Db       4411 TGAAGATCAGGGATTCACTTTTGATGCAGGGCCTACCGTTATTACCGATCCCAGCGCCAT 4470

Qy       3873 CGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGT 3932
               |||||  | || || |||||| ||||||  | | |  |||||||| || ||| |||||||
Db       4471 TGAAGAACTCTTTACCCTGGCGGGAAAACAGTTAAAAGATTACGTTGAACTGCTGCCGGT 4530

Qy       3933 AACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCA 3992
                |||| || |||||||||| | |||||    || ||     |  |||||||| ||  ||||
Db       4531 TGCGCCGTTTTATCGCCTGTGTTGGGAGTCAGGAAAGGTTTTCAATTACGATAACGATCA 4590

Qy       3993 GCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCA 4052
              | | | ||| |||  |||||      ||| ||||| |  ||||| ||||||||||||||||
Db       4591 GGCACAGCTTGAGGCGCAGATTCAGCAGTTTAATCCACGCGATGTTGAAGGCTATCGTCA 4650

Qy       4053 ATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCC 4112
               ||||| | |||||||||||  || |||| ||| || |||||||| || ||||| |||||
Db       4651 GTTTCTGGACTATTCACGTGCGGTGTTTAAAGAAGGCTATCTGAAGCTTGGCACCGTGCC 4710

Qy       4113 GTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATG 4172
               ||  ||  | |  | || ||||| |||| |||||| ||  |||     ||||| |||||
Db       4711 TTTCTTGTCGTTCAGGGATATGCTTCGCGCCGCGCCCCAACTGGCGAAACTGCAGGCATG 4770
```

```
Qy       4173 GCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTT 4232
              ||| | ||| ||||| |  || || |  |  ||| |||| || ||||| ||||||||||
Db       4771 GCGAACCGTTTACAGTAAAGTTGCCAGCTACATTGAGGATGAACATCTTCGTCAGGCGTT 4830

Qy       4233 TTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTT 4292
              ||| || |||||  || | ||||||| ||||| || || || || || || ||||||  |
Db       4831 TTCTTTTCACTCGCTGTTAGTGGGCGGCAATCCCTTCGCCACCTCCTCCATTTATACGCT 4890

Qy       4293 AATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCT 4352
              ||| ||||| || || || |||||||||| |||||||| ||||| || |||||| ||  |
Db       4891 AATACATGCACTCGAACGCGAATGGGGCGTCTGGTTTCCACGCGGTGGTACCGGGGCATT 4950

Qy       4353 GGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGA 4412
               ||| | || |||      ||||||  ||||  |||| |||||  || |||||||||  |
Db       4951 AGTGAAAGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTGGTACTGAATGCGAA 5010

Qy       4413 AGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACG 4472
               || |||||  |||||||||  ||  ||  ||||   | ||| || |||||||| ||||||
Db       5011 GGTCAGCCACATGGAAACCACGGGGGATACCATTGAAGCCGTGCATTTAGAGGACGGACG 5070

Qy       4473 ACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACT 4532
                | |||    ||   |||||| || ||||||||| || |||||| || ||||||   | | ||
Db       5071 CAGATTCCCGACCCGGGCTGTCGCTTCCAATGCGGATGTGGTTCACACCTATCGCGACCT 5130

Qy       4533 GCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAG 4592
              | |  | || |||||    ||  |||| |   ||| |    ||| |   ||||||||||||
Db       5131 GTTAAGTCAGCATCCTGCCGCCGTGAAGCAGTCGAAAAAACTGCAAACTAAGCGCATGAG 5190

Qy       4593 CAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCA 4652
              ||||||  ||||||| ||||||||||| |||||||| |  || || |||||| || |||||
Db       5191 CAACTCATTGTTTGTTCTCTATTTTGGTCTGAATCACCATCACGATCAGCTTGCCCACCA 5250

Qy       4653 CACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCA 4712
               || || ||||| ||||| || || ||||||| ||||  |||| |||||||||      |
Db       5251 TACGGTTTGTTTCGGCCCTCGCTACCGTGAGCTGATTCATGAAATTTTCAATCATGACGG 5310

Qy       4713 GCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGC 4772
               || |||||||||||| |||||||||||||||||| || ||   ||  |||  |||| |||||
Db       5311 ACTCGCAGACGATTTCTCACTTTACCTGCACGCACCTTGTGTCACGGATTCGTCACTGGC 5370

Qy       4773 ACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGA 4832
              ||||   || |||||||| |  |||||||||||| || ||||| ||| | ||||| ||  |
Db       5371 ACCGGAAGGTTGCGGCAGTTACTATGTGTTAGCTCCCGTGCCACATTTAGGCACAGCGAA 5430

Qy       4833 CATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCA 4892
              | ||||||||      |||| | |||||  ||||||||| |||||||  || || |||||
Db       5431 CCTCGACTGGACGGTTGAAGGGCCGCGACTGCGCGATCGTATTTTTGAGTACCTTGAGCA 5490

Qy       4893 GCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGA 4952
               || |||||||| || || |||   || || || || ||  ||||||||||||||||| ||
Db       5491 ACATTACATGCCTGGCTTGCGTAGCCAGTTGGTTACGCAGCGAATGTTTACGCCGTTCGA 5550

Qy       4953 TTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGAC 5012
              ||||||||||  |||  |||||| |||| ||||| || ||||||  | |||||  |  || | ||
Db       5551 TTTTCGCGACCAGCTCAATGCCTATCAAGGCTCAGCCTTTTCTGTTGAGCCCGTTCTCAC 5610

Qy       5013 GCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGT 5072
               || |||||||||||||| |||| ||||||||||      ||||  |||| || |||||
Db       5611 ACAGAGCGCCTGGTTCCGACCGCACAACCGCGATAAAACCATCAATAATCTTTACCTGGT 5670

Qy       5073 GGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCAC 5132
               ||||| || || ||||| |||||||| | ||||||||||||||| |||| ||||| ||
Db       5671 AGGTGCAGGCACCCATCCCGGCGCGGGCATTCCCGGCGTGATCGGCTCGGCTAAGGCAAC 5730

Qy       5133 CGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTA 5192
               ||  |  | |||||||||||   |      |||||   |  | | |||||  | || |
Db       5731 TGCAGGTTTAATGCTGGAGGA---CCTGATTTGAATAATCCGTCGTTACTCAATCATGCG 5787

Qy       5193 ACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCA 5252
                  |||| |||||| || |||||||| |||||| || || ||| | ||| | ||||| |||
Db       5788 GTCGAAACGATGGCTGTTGGCTCGAAAAGTTTTGCGACTGCCTCAAAGTTATTTGACGCA 5847
```

```
Qy          5253 CCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATT 5312
                    || || |||||    || ||||| ||  | ||||| ||| || || || || ||||||
Db          5848 AAAACCCGGCGCAGTGTACTTATGCTCTACTCCTGGTGCCGCCATTGTGACGACGTGATT 5907

Qy          5313 GATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGT 5372
                 || |  ||||| |||||         |   ||||   |  |  ||    |  ||  ||| 
Db          5908 GACGATCAAACCCTGGGTTTTCAAACCGATCAGCCGTCATTACAAACGCCTGAGCAACGC 5967

Qy          5373 ATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCG 5432
                  ||  |||  || | ||  |||| ||||  |||||   |||  |||| ||| | ||||||
Db          5968 CTGATGCAGTTGGAGATGAAAACGCGCCAGGCCTATGCCGGATCGCAGATGCACGAACCG 6027

Qy          5433 GCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTT 5492
                 ||||||   |||||||||||||||||||||    ||  |  |  |||        ||||||
Db          6028 GCGTTTGCTGCGTTTCAGGAAGTGGCGATGGCCCATGATATTGCGCCCGCTTACGCGTTT 6087

Qy          5493 GATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGAC 5552
                 ||||||||||||||||| || |||||||  ||  | |   | |||       || || || 
Db          6088 GATCATCTGGAAGGCTTTGCCATGGATGTGCGTGAGGCGGAGTACATTCAGTTGGACGAT 6147

Qy          5553 ACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATG 5612
                 ||  | || || |||||||||||| || |||||||||||| |||||||||||||   | |||
Db          6148 ACATTACGGTATTGCTATCACGTTGCAGGCGTGGTCGGCTTGATGATGGCGCAGATTATG 6207

Qy          5613 GGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTC 5672
                 ||||| |||||  | ||    || || | |||||| ||  | || ||||| || ||| | 
Db          6208 GGCGTACGCGATAACGCTACACTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAGTTG 6267

Qy          5673 ACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAA 5732
                 || ||||| || ||||| ||||| ||||||||  |    ||||| || ||||||||   
Db          6268 ACCAACATCGCCCGCGATATTGTGGAAGATGCGCATGTGGGTCGTTGTTATCTGCCGGCG 6327

Qy          5733 TCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCA 5792
                    ||||| ||  | |  ||| |       || |      | |||||| || | ||    
Db          6328 AGTTGGCTGGAAAATGAAGGACTGAATAAAGAAAATTATGCGGCACCGGAAAACCGCCAG 6387

Qy          5793 GCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCG 5852
                 |||||   |    | ||    ||| | |||  ||| || ||||| || |||   |  || 
Db          6388 GCGCTGAGCCGCATCGCCCGACGTCTGGTGCAGGAAGCCGAACCTTACTATTTGTGCGCT 6447

Qy          5853 CGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGC 5912
                     | ||  |  | ||  | ||  ||||||| || ||||||||||| || ||       
Db          6448 ACGGCGGGACTGGCCGGGCTGCCCTTGCGCTCAGCCTGGGCCATCGCCACAGCGAAGCAG 6507

Qy          5913 GTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGG 5972
                 ||||||||  ||||||| || || ||| | || || ||     | ||||||||    |||
Db          6508 GTTTATCGTAAAATTGGGGTGAAGGTTGAACAGGCAGGAACAGAAGCCTGGGACCATCGG 6567

Qy          5973 CAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATC 6032
                 |||   ||||  | |   ||||||| |  |||| ||| |||  |   | |||  ||| || 
Db          6568 CAGTCAACCACGACACCCGAAAAATTATCGCTTCTGTTGATGGCATCCGGTCAGGCCATT 6627

Qy          6033 ACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGA 6092
                 ||||| ||  ||   | || |   || || || ||   ||| |||||||| ||       
Db          6628 ACTTCCCGGATGCGGCCTCATCCGCCTCGCCCTGCCCATCTCTGGCAGCGCCCGA----- 6682

Qy          6093 TTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGA 6152
                  ||||      || || |  |  ||||| || || || ||    ||||| ||||| ||||
Db          6683 -TTTAGCCGAATGTCGTTCTCTTAGCGTCGCCTGAAGTTTTGATAGCGGCGGCGCATAGA 6741

Qy          6153 GGAAACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCA 6212
                 | |||||||| || || ||||||||    || |  || ||||||||||||||||| ||||
Db          6742 GAAAACCAAAAGAAACACAGCCTTCTTTGCCTCTTACGGCATGATGCATGCGGTGGGCCA 6801

Qy          6213 TGTATAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCA 6272
                 | || || ||||| || |||||||||||||| ||||   | |||||||| || |||||||
Db          6802 TATACAACCGCTTCAGGTAGCCTTTGCGCGGAATATCCGGCAACGGCCAACGCTGATGCA 6861

Qy          6273 CCAGGCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCC 6332
                 |||| || ||||| |||||||| ||||||   || || | |||||| || ||||||||||
Db          6862 CCAGCCCGTCGTGAACCATGAAATAGAGCAATCCATAGGCCGTCATGCCCGCACCAATCC 6921

Qy          6333 ACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCA 6374
```

```
                 |||| ||||||||||| ||      || |||||||| ||||
Db          6922 ACTGAAGCGGCCACATTCCCGTACTGCCCACATAAATAAGCA 6963

RESULT 6
ERWCRTS
LOCUS       ERWCRTS                12753 bp    DNA     linear   BCT 11-APR-2001
DEFINITION  Pantoea agglomerans pyrophosphate synthase (crtE), zeaxanthin
            glucosyl transferase (crtX), and lycopene cyclase (crtY) genes,
            complete cds; phytoene dehydrogenase (crtI) gene, partial cds;
            phytoene synthase (crtB) and beta-carotene hydroxylase (crtZ)
            genes, complete cds; and unknown genes.
ACCESSION   M87280 M99707
VERSION     M87280.1  GI:148404
KEYWORDS    .
SOURCE      Pantoea agglomerans
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1  (bases 1 to 12753)
  AUTHORS   Hundle,B., Alberti,M., Nievelstein,V., Beyer,P., Kleinig,H.,
            Armstrong,G.A., Burke,D.H. and Hearst,J.E.
  TITLE     Functional assignment of Erwinia herbicola Eho10 carotenoid genes
            expressed in Escherichia coli
  JOURNAL   Mol. Gen. Genet. 245 (4), 406-416 (1994)
  MEDLINE   95107236
   PUBMED   7808389
REFERENCE   2  (bases 1 to 12753)
  AUTHORS   Armstrong,G.A., Hundle,B.S. and Hearst,J.E.
  TITLE     Evolutionary conservation and structural similarities of carotenoid
            biosynthesis gene products from photosynthetic and
            nonphotosynthetic organisms
  JOURNAL   Meth. Enzymol. 214, 297-311 (1993)
  MEDLINE   93225848
   PUBMED   8469144
REFERENCE   3  (bases 1 to 12753)
  AUTHORS   Hundle,B.S., O'Brien,D.A., Beyer,P., Kleinig,H. and Hearst,J.E.
  TITLE     In vitro expression and activity of lycopene cyclase and
            beta-carotene hydroxylase from Erwinia herbicola
  JOURNAL   FEBS Lett. 315 (3), 329-334 (1993)
  MEDLINE   93138098
   PUBMED   8422926
REFERENCE   4  (bases 1 to 12753)
  AUTHORS   Armstrong,G.A., Alberti,M. and Hearst,J.E.
  TITLE     Conserved enzymes mediate the early reactions of carotenoid
            biosynthesis in nonphotosynthetic and photosynthetic prokaryotes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9975-9979 (1990)
  MEDLINE   91088634
   PUBMED   2263648
REFERENCE   5  (sites)
  AUTHORS   Hundle,B.S., O'Brien,D.A., Alberti,M., Beyer,P. and Hearst,J.E.
  TITLE     Functional expression of zeaxanthin glucosyltransferase from
            Erwinia herbicola and a proposed uridine diphosphate binding site
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (19), 9321-9325 (1992)
  MEDLINE   93028456
   PUBMED   1409639
FEATURES             Location/Qualifiers
     source          1..12753
                     /organism="Pantoea agglomerans"
                     /mol_type="genomic DNA"
                     /strain="Eho10"
                     /db_xref="taxon:549"
                     /tissue_lib="ATCC 39368"
     CDS             complement(<1..1588)
                     /note="ORF2"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA64974.1"
                     /db_xref="GI:551854"
                     /translation="MSELIEHANAIEWRFERQILRLEPWGKNSLRVRATCAPEFTSAL
                     HALLTPESCQAEILRDAETLTLRNGNISAVLNLKGQLAFYNQRGELLLEEMWRQRSTV
                     GIGASEKSQDKYVSALKLDGREFKPLPGGKYQLTVRFEARQGEKLYGMGQYQQPWLDL
                     KGCSLELAQRNSQASVPFVQSSLGYGLLWNNPAIGEASFAKNHSQWTSRVTQEMDYWI
                     TAGDSVMEITRQYAKATGTPPAAPEFISGLWQCKLRYRTQQEVLDVAREYHRRNLPLS
                     VMVIDFFHWPNQGTWCFDPVDWPDPRAMVEELESMGIRLMVSVWPTVEARSPLFPQMK
                     AKGWLVTSDRGVQVNLDFMGNTTFFDATHPQARQFVWDTVKKNYFDYGIKLFWLDEAE
                     PEYRAYDFDNYRYHAGPVLEVGNRYPRDFAQGFYDGLRAEGENEIVNLVRCSWAGSQR
                     YGVLAWSGDVHSSFHSFRNQLAAGLNMGLAGIPWWTTDIGGFQGGNIHDPAFHELLIR
                     WFQWAVFCPVLRMHGYREPRI"
     CDS             complement(1596. .2414)
                     /note="ORF3"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA64975.1"
                     /db_xref="GI:148406"
                     /translation="MLTMNNLLMVLVQALRQHRWLCLFGVAYIFSSVGNGLTQTLILG
                     QLLRWHAPPSTLTLTYMLATLPGFIGSYLGAMMYFVTYYLGSASYFMWMLAAHILGKA
                     AGSLLAKRLTRNFNKVQIFGYCAVLAGVLSIALFFAPKSVFVLVPLTFIISTLYQATT
                     TLMWVMMADVADYGEWSQGKRMDGIIFSTFLAVLKLGMALSGAIVGWTLGFSGYVANA
                     PEQTSLAMHCIVALFTVVPGLLSLAAFATLRWYKLDDQTMQEINLAKMQTIVKG"
     CDS             3140. .3451
                     /note="ORF4"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA64976.1"
                     /db_xref="GI:148407"
                     /translation="MALRSLGTTGLHIPPLVFGGNVFGWTVDEQQSFSLLDALLERGL
                     NAIDTADVYSAWAPGNKGGESETILGKWFAAHPGAREKITLFTKVGSDLKAQRLMLRT
                     M"
     gene            3535. .4458
                     /gene="crtE"
     CDS             3535. .4458
                     /gene="crtE"
                     /citation=[4]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="pyrophosphate synthase"
                     /protein_id="AAA64977.1"
                     /db_xref="GI:148408"
                     /translation="MVSGSKAGVSPHREIEVMRQSIDDHLAGLLPETDSQDIVSLAMR
                     EGVMAPGKRIRPLLMLLAARDLRYQGSMPTLLDLACAVELTHTASLMLDDMPCMDNAE
                     LRRGQPTTHKKFGESVAILASVGLLSKAFGLIAATGDLPGERRAQAVNELSTAVGVQG
                     LVLGQFRDLNDAALDRTPDAILSTNHLKTGILFSAMLQIVAIASASSPSTRETLHAFA
                     LDFGQAFQLLDDLRDDHPETGKDRNKDAGKSTLVNRLGADAARQKLREHIDSADKHLT
                     FACPQGGAIRQFMHLWFGHHLADWSPVMKIA"
     CDS             4521. .5564
                     /note="gene not found in Erwinia uredovora crt gene
                     cluster; ORF6"
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="unknown"
                     /protein_id="AAA64978.1"
                     /db_xref="GI:148409"
                     /translation="MKDERLVQRKNDHLDIVLDPRRAVTQASAGFERWRFTHCALPEL
                     NFSDITLETTFLNRQLQAPLLISSMTGGVERSRHINRHLAEAAQVLKIAMGVGSQRVA
                     IESDAGLGLDKTLRQLAPDVPLLANLGAAQLTGRKGIDYARRAVEMIEADALIVHLNP
                     LQEALQPGGDRDWRGRLAAIETLVRELPVPLVVKEVGAGISRTVAGQLIDAGVTVIDV
                     AGAGGTSWAAVEGERAATEQQRSVANVFADWGIPTAEALVDIAEAWPQMPLIASGGIK
                     NGVDAAKALRLGACMVGQAAAVLGSAGVSTEKVIDHFNVIIEQLRVACFCTGSRSLSD
                     LKQADIRYVRDTP"
     gene            5561. .6802
                     /gene="crtX"
     CDS             5561. .6802
                     /gene="crtX"
                     /citation=[3]
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
```

```
                     /product="zeaxanthin glucosyl transferase"
                     /protein_id="AAA64979.1"
                     /db_xref="GI:148410"
                     /translation="MSHFAIVAPPLYSHAVALHALALEMAQRGHRVTFLTGNVASLAE
                     QETERVAFYPLPASVQQAQRNVQQQSNGNLLRLIAAMSSLTDVLCQQLPAILQRLAVD
                     ALIVDEMEPAGSLVAEALGLPFISIACALPVNREPGLPLPVMPFHYAEDKRALRRFQV
                     SERIYDALMYPHGQTILRHAQRFGLPERRRLDECLSPLAQISQSVPALDFPRRALPNC
                     FHYVGALRYQPPPQVERSPRSTPRIFASLGTLQGHRLRLFQKIARACASVGAEVTIAH
                     CDGLTPAQADSLYACGATEVVSFVDQPRYVAEANLVITHGGLNTVLDALAAATPVLAV
                     PLSFDQPAVAARLVYNGLGRRVSRFARQQTLADEIAQLLGDETLHQRLATARQQLNDA
                     GGTPRAATLIEQAIAGSESVS"
     gene            6799. .7959
                     /gene="crtY"
     CDS             6799. .7959
                     /gene="crtY"
                     /citation=[3]
                     /citation=[2]
                     /citation=[1]
                     /codon_start=1
                     /transl_table=11
                     /product="lycopene cyclase"
                     /protein_id="AAA64980.1"
                     /db_xref="GI:148411"
                     /translation="MRDLILVGGGLANGLIAWRLRQRYPQLNLLLIEAGEQPGGNHTW
                     SFHEDDLTPGQHAWLAPLVAHAWPGYEVQFPDLRRRLARGYYSITSERFAEALHQALG
                     ENIWLNCSVSEVLPNSVRLANGEALLAGAVIDGRGVTASSAMQTGYQLFLGQQWRLTQ
                     PHGLTVPILMDATVAQQQGYRFVYTLPLSADTLLIEDTRYANVPQRDDNALRQTVTDY
                     AHSKGWQLAQLEREETGCLPITLAGDIQALWADAPGVPRSGMRAGLFHPTTGYSLPLA
                     VALADAIADSPRLGSVPLYQLTRQFAERHWRRQGFFRLLNRMLFLAGREENRWRVMQR
                     FYGLPEPTVERFYAGRLSLFDKARILTGKPPVPLGEAWRAALNHFPDRRDKG"
     gene            7956. .>9414
                     /gene="crtI"
     CDS             7956. .>9414
                     /gene="crtI"
                     /note="putative; same bases, 11. .1489, Genbank M38423"
                     /citation=[4]
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene dehydrogenase"
                     /protein_id="AAA64981.1"
                     /db_xref="GI:551855"
                     /translation="MKKTVVIGAGFGGLALAIRLQAAGIPTVLLEQRDKPGGRAYVWH
                     DQGFTFDAGPTVITDPTALEALFTLAGRRMEDYVRLLPVKPFYRLCWESGKTLDYAND
                     SAELEAQITQFNPRDVEGYRRFLAYSQAVFQEGYLRLGSVPFLSFRDMLRAGPQLLKL
                     QAWQSVYQSVSRFIEDEHLRQAFSFHSLLVGGNPFTTSSIYTLIHALEREWGVWFPEG
                     GTGALVNGMVKLFTDLGGEIELNARVEELVVADNRVSQVRLADGRIFDTDAVASNADV
                     VNTYKKLLGHHPVGQKRAAALERKSMSNSLFVLYFGLNQPHSQLAHHTICFGPRYREL
                     IDEIFTGSALADDFSLYLHSPCVTDPSLAPPGCASFYVLAPVPHLGNAPLDWAQEGPK
                     LRDRIFDYLEERYMPGLRSQLVTQRIFTPADFHDTLDAHLGSAFSIEPLLTQSAWFRP
                     HNRDSDIANLYLVGAGTHPGAGIPGVVASAKATASL"
     gene            9431. .10360
                     /gene="crtB"
     CDS             9431. .10360
                     /gene="crtB"
                     /note="same bases, 1486. .2415, Genbank M38423 but now
                     function has been reassigned"
                     /citation=[4]
                     /citation=[2]
                     /codon_start=1

  Query Match             30.5%;  Score 2135.2;  DB 1;  Length 12753;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 3263;  Conservative    0;  Mismatches 1765;  Indels   39;  Gaps    4;

Qy       1479 CTGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCT 1538
              ||| | ||  |||| |  |   |  | |||   |||||| |||| ||||   |  ||||
Db       5780 CTGCTGCGGCTGATTGCGGCCATGTCATCCCTGACCGATGTGCTCTGCCAGCAGTTGCCC 5839

Qy       1539 GCGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGC 1598
              ||  | ||  |     ||    | || |   |||| | ||| || |||||  | || ||
Db       5840 GCTATTCTACAGCGGCTGGCGGTGGACGCGCTGATTGTCGATGAGATGGAGCCCGCCGGA 5899
```

```
Qy       1599 GGATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTC 1658
Db       5900 AGCCTGGTCGCCGAGGCGCTGGGACTACCATTTATCTCTATTGCCTGCGCGCTGCCGGTC 5959
Qy       1659 AATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAA 1718
Db       5960 AACCGCGAGCCGGGTCTGCCGCTGCCGGTGATGCCGTTTCACTACGCCGAGGATAAGAGA 6019
Qy       1719 GCGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGC 1778
Db       6020 GCCCTGCGGCGTTTTCAGGTCAGCGAACGGATCTACGATGCGCTGATGTACCCGCACGGG 6079
Qy       1779 GACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAG 1838
Db       6080 CAGACGATCCTGCGCCACGCCCAGCGCTTTGGTTTGCCGGAGCGCAGGCGTCTCGACGAG 6139
Qy       1839 TGCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTCCACGTCAG 1898
Db       6140 TGTCTCTCGCCGCTGGCGCAGATTAGCCAGTCCGTTCCGGCCCTCGACTTCCCACGCCGG 6199
Qy       1899 CAACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCG 1958
Db       6200 GCGCTGCCGAACTGTTTTCACTACGTGGGAGCACTGCGC---------TATCAGCCGCCG 6250
Qy       1959 CTCCATGCGCCCTGGCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTG 2018
Db       6251 CCGCAGGTAGAACGCTCGCCACGCAGCACGCCGCGGATCTTTGCCTCGCTGGGCACCCTC 6310
Qy       2019 CAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTA 2078
Db       6311 CAGGGCCACCGTCTACGCCTGTTTCAGAAGATCGCCCGCGCCTGTGCCAGCGTGGGGGCG 6370
Qy       2079 TCGCTGGTGATCGCCCATTGTGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTC 2138
Db       6371 GAGGTGACCATTGCCCACTGCGATGGCCTGACGCCCGCCCAGGCCGACTCGCTCTACGCC 6430
Qy       2139 GCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAG 2198
Db       6431 TGCGGCGCGACGGAGGTGGTCAGCTTTGTCGACCAGCCGCGCTACGTTGCCGAGGCTAAT 6490
Qy       2199 CTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCG 2258
Db       6491 CTGGTGATCACCCACGGCGGTCTCAATACCGTACTGGATGCGCTGGCTGCCGCGACGCCG 6550
Qy       2259 ATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCAT 2318
Db       6551 GTGCTGGCGGTGCCACTCTCTTTCGACCAGCCCGCCGTGGCTGCCCGGCTGGTCTATAAC 6610
Qy       2319 GACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAA 2378
Db       6611 GGGCTGGGTCGCCGGGTATCGCGCTTTGCCAGACAGCAGACGCTGGCGGATGAGATTGCC 6670
Qy       2379 CAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAG 2438
Db       6671 CAACTGCTGGGGGATGAGACGCTGCATCAGCGTCTGGCGACGGCCCGCCAGCAGCTTAAC 6730
Qy       2439 CGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAA 2498
Db       6731 GACGCCGGGGGCACGCCCCGTGCGGCGACCCTGATTGAACAGGCCAT------------- 6777
Qy       2499 GTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTG 2558
Db       6778 -----------AGCAGGGAGTGAGAGCGTATCGTGAGGGATCTGATTTTAGTCGGCGGCG 6826
Qy       2559 GACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCC 2618
Db       6827 GCCTGGCCAACGGGCTGATCGCCTGGCGTCTGCGCCAGCGCTACCCGCAGCTTAACCTGC 6886
Qy       2619 TGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCG 2678
Db       6887 TGCTGATCGAGGCCGGGGAGCAGCCCGGCGGGAACCATACCTGGTCATTCCATGAAGACG 6946
Qy       2679 ATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTT 2738
```

```
Db       6947 ATCTGACTCCCGGGCAGCACGCCTGGCTGGCCCCGCTGGTGGCCCACGCCTGGCCGGGCT 7006
Qy       2739 ATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCAT 2798
Db       7007 ATGAGGTGCAGTTTCCCGATCTTCGCCGTCGCCTCGCGCGGCTACTACTCCATTACCT 7066
Qy       2799 CAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAG 2858
Db       7067 CAGAGCGCTTTGCCGAGGCCCTGCATCAGGCGCTGGGGGAGAACATCTGGCTAAACTGTT 7126
Qy       2859 CCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGC 2918
Db       7127 CGGTGAGCGAGGTGTTACCCAATAGCGTGCGCCTTGCCAACGGTGAGGCGCTGCTTGCCG 7186
Qy       2919 AAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGG 2978
Db       7187 GAGCGGTGATTGACGGACGCGGCGTGACCGCCAGTTCGGCGATGCAAACCGGCTATCAGC 7246
Qy       2979 TGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGA 3038
Db       7247 TCTTTCTTGGTCAGCAGTGGCGGCTGACACAGCCCCACGGCCTGACCGTACCGATCCTGA 7306
Qy       3039 TGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCG 3098
Db       7307 TGGATGCCACGGTGGCGCAGCAGCAGGGCTATCGCTTTGTCTACACGCTGCCGCTCTCCG 7366
Qy       3099 CCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACA 3158
Db       7367 CCGACACGCTGCTGATCGAGGATACGCGCTACGCCAATGTCCCGCAGCGTGATGATAATG 7426
Qy       3159 CCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGC 3218
Db       7427 CCCTACGCCAGACGGTTACCGACTATGCTCACAGCAAAGGGTGGCAGCTGGCCCAGCTTG 7486
Qy       3219 TGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGC 3278
Db       7487 AACGCGAGGAGACCGGCTGTCTGCCGATTACCCTGGCGGGTGACATCCAGGCTCTGTGGG 7546
Qy       3279 AACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCG 3338
Db       7547 CCGATGCGCCGGGCGTGCCGCGCTCGGGAATGCGGGCTGGGCTATTTCACCCTACCACTG 7606
Qy       3339 GTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCG 3398
Db       7607 GCTATTCGCTGCCGCTGGCGGTGGCCCTTGCCGACGCGATTGCCGACAGCCCGCGGCTGG 7666
Qy       3399 ATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGC 3458
Db       7667 GCAGCGTTCCGCTCTATCAGCTCACCCGGCAGTTTGCCGAACGCCACTGGCGCAGGCAGG 7726
Qy       3459 GATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGC 3518
Db       7727 GATTCTTCCGCCTGCTGAACCGGATGCTTTTCCTGGCCGGGCGCGAGGAGAACCGCTGGC 7786
Qy       3519 GCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGC 3578
Db       7787 GGGTGATGCAGCGCTTTTATGGGCTGCCGGAGCCCACCGTAGAGCGCTTTTACGCCGGTC 7846
Qy       3579 AACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTG 3638
Db       7847 GGCTCTCTCTCTTTGATAAGGCCCGCATTTTGACGGGCAAGCCACCGGTTCCGCTGGGCG 7906
Qy       3639 AAGCGCTGCGCGCGCTGTTGA---ATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCAC 3695
Db       7907 AAGCCTGGCGGGCGGCGCTGAACCATTTTCCTGACAGACGAGATAAAGGATGAAAAAAAC 7966
Qy       3696 TTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGG 3755
Db       7967 CGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTCGCCTGCAGGCGGCAGG 8026
Qy       3756 CATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGA 3815
```

```
Db      8027 GATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTCGGGCCTACGTCTGGCA 8086

Qy      3816 GGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGA 3875
              |||   ||||||||||| || ||||| || |||||||||||||||||| | |||  | ||
Db      8087 TGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCGATCCTACCGCGCTTGA 8146

Qy      3876 AGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAAC 3935
              | | ||||||| ||||| || |      |    |||||||||  |||| ||||||||| 
Db      8147 GGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCAGGCTGCTGCCGGTAAA 8206

Qy      3936 GCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCC 3995
              |||||||| || || ||||||||    || ||    || || || |  ||  |  |     |
Db      8207 ACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACTATGCTAACGACAGCGC 8266

Qy      3996 GCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATT 4055
                ||| |||  ||||||  ||  |||||| || |  || || || ||||| || |  ||
Db      8267 CGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCGAGGGCTACCGGCGCTT 8326

Qy      4056 TCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTT 4115
             ||| || || || |  |  |||||    ||||| ||| ||   |||||||  ||||||||
Db      8327 TCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCCTCGGCAGCGTGCCGTT 8386

Qy      4116 TCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCG 4175
              ||     | || ||||||||||||| || ||||||| ||      || || || |||| 
Db      8387 CCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTAAGCTCCAGGCGTGGCA 8446

Qy      4176 CAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTC 4235
              |||||||||     |||  ||   |||||| |||| || |||||||| |||| || ||
Db      8447 GAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATCTGCGGCAGGCCTTCTC 8506

Qy      4236 CTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAAT 4295
              ||||||||  ||||||| ||||| || || ||  | || || || ||||| ||| | ||
Db      8507 GTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGTCCATCTACACCCTGAT 8566

Qy      4296 TCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGT 4355
              || || || ||||| || ||||| || ||||| ||    ||||||||||| ||||||||
Db      8567 CCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCGGCACCGGGGCGCTGGT 8626

Qy      4356 GCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGT 4415
             | | ||||||| |   |||||    ||  |||||||| ||| |   ||| || |||   ||
Db      8627 GAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCGAACTCAACGCCCGGGT 8686

Qy      4416 GAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACG 4475
                  ||||||       ||  || ||| | |||   || | | | | ||  || ||   
Db      8687 CGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGCTGGCGGATGGTCGGAT 8746

Qy      4476 CTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCT 4535
             ||| ||  ||| ||| || ||||| || || ||||||||| | |||||  | || |||||
Db      8747 CTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACACCTATAAAAAGCTGCT 8806

Qy      4536 TCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAA 4595
               ||||||||||| |||    ||| || ||| || || ||||| |||| ||| ||||||||||
Db      8807 CGGCCACCATCCGGTGGGGCAGAAGCGGGCGGCAGCGCTGGAGCGCAAGAGCATGAGCAA 8866

Qy      4596 CTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACAC 4655
             |||||||||||| ||||| || |||||||| ||||| |||   ||||| ||||||| ||
Db      8867 CTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCCAGCTGGCGCACCATAC 8926

Qy      4656 CGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCT 4715
             | |||||||||| || || || || ||| |||||||| ||||| || || | | |||||  |||
Db      8927 CATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCTTTACCGGCAGCGCGCT 8986

Qy      4716 GGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACC 4775
             ||| || || || || || |||||||| |||||||||   | |||||| ||||| || ||
Db      8987 GGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCGATCCCTCGCTCGCGCC 9046

Qy      4776 GCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACAT 4835
              |||||||||| |||||| || ||| | || ||||||||||||||| ||||  |||     |
Db      9047 TCCCGGCTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATCTTGGCAACGCGCCGCT 9106

Qy      4836 CGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCA 4895
              ||||||   ||||| || |||    |||||| || || ||||  || || ||  |||  
Db      9107 GGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTGACTACCTTGAAGAGCG 9166

Qy      4896 CTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTT 4955
             ||| ||||| ||  | |||   ||  | ||||| ||  | || ||||| |||   || ||
Db      9167 CTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCTTTACCCCGGCAGACTT 9226

Qy      4956 TCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCA 5015
              | ||||||||||| |||| ||||  || ||||| || || |  | |||||| |  |||| ||
Db      9227 CCACGACACGCTGGATGCGCATCTGGGATCGGCCTTCTCCATCGAGCCGCTGCTGACCCA 9286

Qy      5016 AAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGG 5075
             ||||||||||||||||||||| ||||||||   ||| ||   ||| ||||| ||||||||
Db      9287 AAGCGCCTGGTTCCGCCCGCACAACCGCGACAGCGACATTGCCAACCTCTACCTGGTGGG 9346

Qy      5076 TGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGC 5135
              || ||||| || || || |||||| | || |||||  | |  ||||| || ||||||||
Db      9347 CGCAGGTACTCACCCTGGGGCGGGCATTCCTGGCGTAGTGGCCTCGGCGAAAGCCACCGC 9406

Qy      5136 CAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACG 5195
             ||| |||||| | |||||||      |||||  | || ||| |  | ||||| || |  |||
Db      9407 CAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCTGCTTGACCACGCCACG 9463

Qy      5196 CAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCG 5255
             || ||||||||    ||||||||| ||||| |||||||| || || |||||||| ||  |  | 
Db      9464 CAGACCATGGCCAACGGCTCGAAAAGTTTTGCCACCGCTGCGAAGCTGTTCGACCCGGCC 9523

Qy      5256 ACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGAT 5315
             || ||||| |||  |||||||||| ||  | ||||| || ||||||||||||| || ||||| 
Db      9524 ACCCGCCGTAGCGTGCTGATGCTCTACACCTGGTGCCGCCACTGCGATGACGTCATTGAC 9583

Qy      5316 GGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATG 5375
             |  || || |  |||   | | ||  |      || |  || || ||       ||  ||
Db      9584 GACCAGACCCACGGCTTCGCCAGCGAGGCCGCGGCGGAGGAGGAGGCCACCCAGCGCCTG 9643

Qy      5376 CAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCG 5435
                |  ||||  |     ||||     || |     || ||  | ||| |  || ||||| 
Db      9644 GCCCGGCTGCGCACGCTGACCCTGGCGGCGTTTGAAGGGGCCGAGATGCAGGATCCGGCC 9703

Qy      5436 TTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGAT 5495
             ||    || |||||||| ||||||| | | |  |||    |  |||  |   ||||| | |||
Db      9704 TTCGCTGCCTTTCAGGAGGTGGCGCTGACCCACGGTATTACGCCCCGCATGGCGCTCGAT 9763

Qy      5496 CATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACG 5555
             || || || ||||| || |||| |      |    |  | |  | |||  || |||
Db      9764 CACCTCGACGGCTTTGCGATGGACGTGGCTCAGACCCGCTATGTCACCTTTGAGGATACG 9823

Qy      5556 CTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGC 5615
             ||||| |||||||||||||| ||||||||||| ||| ||||||||||  | || ||||||
Db      9824 CTGCGCTACTGCTATCACGTGGCGGGCGTGGTGGGTCTGATGATGGCCAGGGTGATGGGC 9883

Qy      5616 GTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACT 5675
             ||||| || ||   |||||| |||| |||||||||| | || ||||| |||||||| || 
Db      9884 GTGCGGGATGAGCGGGTGCTGGATCGCGCCTGCGATCTGGGGCTGGCCTTCCAGCTGACG 9943

Qy      5676 AACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCC 5735
             || || || || || ||| | || |||||  |  | ||  ||||||||||||||       
Db      9944 AATATCGCCCGGGATATTATTGACGATGCGGCTATTGACCGCTGCTATCTGCCCGCCGAG 10003

Qy      5736 TGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCG 5795
             |||||  |  | || ||  |   | | || |      | || | | | |  |||| || |||
Db     10004 TGGCTGCAGGATGCCGGGCTGACCCCGGAGAACTATGCCGCGCGGGAGAATCGGGCCGCG 10063

Qy      5796 CTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGA 5855
             || ||     |||| | |||  |  | |  |  || || || || ||   |||  | |  
Db     10064 CTGGCGCGGGTGGCGGAGCGGCTTATTGATGCCGCAGAGCCGTACTACATCTCCTCCCAG 10123

Qy      5856 TCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTT 5915
              ||||  |||  | | | |||| |||||  || ||||| |||| || || ||| |||| 
Db     10124 GCCGGGCTACACGATCTGCCGCCGCGCTGCGCCTGGGCGATCGCCACCGCCCGCAGCGTC 10183

Qy      5916 TATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAG 5975
             || || || || ||  | || ||  |    || || |    |||||||||||   || |||
Db     10184 TACCGGGAGATCGGTATTAAGGTAAAAGCGGCGGGAGGCAGCGCCTGGGATCGCCGCCAG 10243
```

```
Qy       5976 CGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACT 6035
              | |||||| |||||||||||| | ||  ||||| ||   |   | ||   ||  ||
Db      10244 CACACCAGCAAAGGTGAAAAAATTGCCATGCTGATGGCGGCACCGGGGCAGGTTATTCGG 10303

Qy       6036 TCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTT 6095
               ||     | |  |       ||||||||||| ||||| |||||||||||   ||
Db      10304 GCGAAGACGACGAGGGTGACGCCGCGTCCGGCCGGTCTTTGGCAGCGTCCCGTTTAGGCG 10363

Qy       6096 TACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGA 6155
                || || ||||| |  |||||  | || || || |      ||    | ||||| |  |
Db      10364 GGCGGCCATGACGTTCACGCAGGATCGCCTGTAGGTCGGCAGGCTTGCGGGCGTAAATAA 10423

Qy       6156 AACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGT 6215
              |||| || || |||||||| || || ||||||||||| || |||| |||||| || | ||
Db      10424 AACCGAAGGAGACGCAGCCCTCCCGGCCGCGCACCGCGTGGTGCAGGCGGTGGGCGACGT 10483

Qy       6216 ATAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCA 6275
              | |  ||||| || |||||   |||||||||  || |||| |||||||| ||||||||||
Db      10484 AGAGCCGCTTCAGGTAGCCCCGGCGCGGGATCCAGTGGAAGGGCCAGCGCTGATGCACCA 10543

Qy       6276 GGCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACT 6335
              | || |||||||||| ||||||||||  ||| ||   |||||| |||   ||||||||||
Db      10544 GACCGTCGTGCACCAGGAAGTAGAGCAGGCCATAGACCGTCATGCCGCAGCCAATCCACT 10603

Qy       6336 GCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACA 6395
              |||| ||||| | |||   |   || ||   ||||||| | || ||||   | |||||||
Db      10604 GCAGGGGCCAAACGCCCGCCGTGCCCACGGCAATCAGCGCGATAGCCACCCCGGCAAACA 10663

Qy       6396 CCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGAT 6455
              |||||||| | |||||||| ||||||||    ||  || ||||   |||||| |||  ||
Db      10664 CCACCGCAAAGAGATCGTTTAGCTCAAATACGCCCTTGCGCGGGGTATGGTGTGACTCAT 10723

Qy       6456 GCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCA 6515
              |||||| ||||||||| |||||||| |||||    || |  | ||| || | |   ||||
Db      10724 GCCAGCGCCATCCCCAGCCGTGCATAATGTAGCGGTGGGTAAACGCGGCGATGCCCTCCA 10783

Qy       6516 TCACCACCACGGTTGCCAACAAGATAA 6542
              || | |  ||| |    |  | ||| |
Db      10784 TCGCAATAACGCTCAAGATGACGATTA 10810


RESULT 7
CQ793208
LOCUS       CQ793208                4624 bp    DNA     linear   PAT 29-MAR-2004
DEFINITION  Sequence 24 from Patent WO2004018694.
ACCESSION   CQ793208
VERSION     CQ793208.1  GI:45824305
KEYWORDS    .
SOURCE      Pantoea ananatis
  ORGANISM  Pantoea ananatis
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1
  AUTHORS   Sauer,M., Flachmann,R., Klebsattel,M. and Schopfer,C.R.
  TITLE     Method for producing ketocarotinoids in genetically modified
            organisms
  JOURNAL   Patent: WO 2004018694-A 24 04-MAR-2004;
            Sungene GmbH & Co. KgAA (DE)
FEATURES             Location/Qualifiers
     source          1..4624
                     /organism="Pantoea ananatis"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:553"
     CDS             128..>1267
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAG24421.1"
                     /db_xref="GI:45824306"
                     /translation="MQPHYDLILVGAGLANGLIALRLQQQQPDMRILLIDAAPQAGGN
                     HTWSFHHDDLTESQHRWIAPLVVHHWPDYQVRFPTRRRKLNSGYFCITSQRFAEVLQR
                     QFGPHLWMDTAVAEVNAESVRLKKGQVIGARAVIDGRGYAANSALSVGFQAFIGQEWR
                     LSHPHGLSSPIIMDATVDQQNGYRFVYSLPLSPTRLLIEDTHYIDNATLDPECARQNI
                     CDYAAQQGWQLQTLLREEQGALPITLSGNADAFWQQRPLACSGLRAGLFHPTTGYSLP
                     LAVAVADRLSALDVFTSASIHHAITHFARERWQQQGFFRMLNRMLFLAGPADSRWRVM
                     QRFYGLPEDLIARFYAGKLTLTDRLRILSGKPPVPVLAALQAIMTT"
     CDS             1288..2766
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAG24422.1"
                     /db_xref="GI:45824307"
                     /translation="MKPTTVIGAGFGGLALAIRLQAAGIPVLLLEQRDKPGGRAYVYE
                     DQGFTFDAGPTVITDPSAIEELFALAGKQLKEYVELLPVTPFYRLCWESGKVFNYDND
                     QTRLEAQIQQFNPRDVEGYRQFLDYSRAVFKEGYLKLGTVPFLSFRDMLRAAPQLAKL
                     QAWRSVYSKVASYIEDEHLRQAFSFHSLLVGGNPFATSSIYTLIHALEREWGVWFPRG
                     GTGALVQGMIKLFQDLGGEVVLNARVSHMETTGNKIEAVHLEDGRRFLTQAVASNADV
                     VHTYRDLLSQHPAAVKQSNKLQTKRMSNSLFVLYFGLNHHHDQLAHHTVCFGPRYREL
                     IDEIFNHDGLAEDFSLYLHAPCVTDSSLAPEGCGSYYVLAPVPHLGTANLDWTVEGPK
                     LRDRIFAYLEQHYMPGLRSQLVTHRMFTPFDFRDQLNAYHGSAFSVEPVLTQSAWFRP
                     HNRDKTITNLYLVGAGTHPGAGIPGVIGSAKATAGLMLEDLI"
     CDS             2802..3692
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAG24423.1"
                     /db_xref="GI:45824308"
                     /translation="MAVGSKSFATASKLFDAKTRRSVLMLYAWCRHCDDVIDDQTLGF
                     QARQPALQTPEQRLMQLEMKTRQAYAGSQMHEPAFAAFQEVAMAHDIAPAYAFDHLEG
                     FAMDVREAQYSQLDDTLRYCYHVAGVVGLMMAQIMGVRDNATLDRACDLGLAFQLTNI
                     ARDIVDDAHAGRCYLPASWLEHEGLNKENYAAPENRQALSRIARRLVQEAEPYYLSAT
                     AGLAGLPLRSAWAIATAKQVYRKIGVKVEQAGQQAWDQRQSTTTPEKLTLLLAASGQA
                     LTSRMRAHPPRPAHLWQRPL"
     iDNA            3631..4158
ORIGIN

  Query Match            27.2%;  Score 1900.6;  DB 6;  Length 4624;
  Best Local Similarity  66.5%;  Pred. No. 0;
  Matches 2787;  Conservative    0;  Mismatches 1389;  Indels   13;  Gaps    4;

Qy       2405 GCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGC 2464
              ||  || ||| |    || ||| |    || |     ||||| ||   |  |   || ||
Db         12 GCAGCGCATGGCGAAAATCCAGACAGCCCTTCGTTTGGCAGGGGGCACCATGGCCGCTGC 71

Qy       2465 CGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATG 2524
              ||| ||| | |||||||   |||| || | |  || |   ||   |  || |     |||
Db         72 CGATATCATTGAGCAGGTTATGTG-CACCGGTCAGCCTGTCTTAAGTGGGAGCGGCTATG 130

Qy       2525 CGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTG 2584
              |   |||| || ||| ||||| | || || |||||||| ||||||||| | || || |||
Db        131 CAACCGCATTATGATCTGATTCTCGTGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTG 190

Qy       2585 CGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCG 2644
              ||||| |  ||   ||| ||  |  ||       || |  | ||   ||   | ||  ||
Db        191 CGTCTTCAGCAGCAGCAACCTGATATGCGTATTTTGCTTATCGACGCCGCACCCCAGGCG 250

Qy       2645 GCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGG 2704
              |  || |||||||| ||||| ||||| |||   ||| | |  |     |||| ||| |||
Db        251 GGCGGGAATCATACGTGGTCATTTCACCACGATGATTTGACTGAGAGCCAACATCGTTGG 310

Qy       2705 CTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGC 2764
               |    |||||| |       |  ||| | |  |||||||| || |||||  | |  ||
Db        311 ATAGCTCCGCTGGTGGTTCATCACTGGCCCGACTATCAGGTACGCTTTCCCACACGCCGT 370

Qy       2765 CGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTAC 2824
              || || ||| || | |  || | ||  ||  | ||      ||| ||      | |  |
Db        371 CGTAAGCTGAACAGCGGCTACTTTTGTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAG 430

Qy       2825 GCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAG 2884
                   | | ||     |  ||||||   | || || ||   | |||| ||  |
Db        431 CGACAGTTTGGCCCGCACTTGTGGATGGATACCGCGGTCGCAGAGGTTAATGCGGAATCT 490

Qy       2885 GTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTG 2944
              ||   | ||    | || |  |   | | ||| |  | ||||||||| || || ||
Db        491 GTTCGGTTGAAAAAGGGTCAGGTTATCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTAT 550
```

```
Qy       2945 CAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTG 3004
Db        551 GCGGCAAATTCAGCACTGAGCGTGGGCTTCCAGGCGTTTATTGGCCAGGAATGGCGATTG 610

Qy       3005 GCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCG 3064
Db        611 AGCCACCCGCATGGTTTATCGTCTCCCATTATCATGGATGCCACGGTCGATCAGCAAAAT 670

Qy       3065 GGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACC 3124
Db        671 GGTTATCGCTTCGTGTACAGCCTGCCGCTCTCGCCGACCAGATTGTTAATTGAAGACACG 730

Qy       3125 CATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTAT 3184
Db        731 CACTATATTGATAATGCGACATTAGATCCTGAATGCGCGCGGCAAAATATTTGCGACTAT 790

Qy       3185 GCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCG 3244
Db        791 GCCGCGCAACAGGGTTGGCAGCTTCAGACACTGCTGCGAGAAGAACAGGGCGCCTTACCC 850

Qy       3245 ATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGC 3304
Db        851 ATTACTCTGTCGGGCAATGCCGACGCATTCTGG---CAGCAGCGCCCCCTGGCCTGTAGT 907

Qy       3305 GGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCG 3364
Db        908 GGATTACGTGCCGGTCTGTTCCATCCTACCACCGGCTATTCACTGCCGCTGGCGGTTGCC 967

Qy       3365 CTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATC 3424
Db        968 GTGGCCGACCGCCTGAGTGCACTTGATGTCTTTACGTCGGCCTCAATTCACCATGCCATT 1027

Qy       3425 GAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATG 3484
Db       1028 ACGCATTTTGCCCGCGAGCGCTGGCAGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATG 1087

Qy       3485 CTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTC 3544
Db       1088 CTGTTTTTAGCCGGACCCGCCGATTCACGCTGGCGGGTTATGCAGCGTTTTTATGGTTTA 1147

Qy       3545 GATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGG 3604
Db       1148 CCTGAAGATTTAATTGCCCGTTTTTATGCGGGAAAACTCACGCTGACCGATCGGCTACGT 1207

Qy       3605 ATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCT 3664
Db       1208 ATTCTGAGCGGCAAGCCGCCTGTTCCGGTATTAGCAGCATTGCAAGCCATTATGACGACT 1267

Qy       3665 GTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCC 3724
Db       1268 CATCGTTAAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCC 1327

Qy       3725 TGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCG 3784
Db       1328 TGGCACTGGCAATTCGTCTACAAGCTGCGGGGATCCCCGTCTTACTGCTTGAACAACGTG 1387

Qy       3785 ACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGAC 3844
Db       1388 ATAAACCCGGCGGTCGGGCTTATGTCTACGAGGATCAGGGGTTTACCTTTGATGCAGGCC 1447

Qy       3845 CCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGC 3904
Db       1448 CGACGGTTATCACCGATCCCAGTGCCATTGAAGAACTGTTTGCACTGGCAGGAAAACAGT 1507

Qy       3905 TCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATG 3964
Db       1508 TAAAAGAGTATGTCGAACTGCTGCCGGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAG 1567

Qy       3965 GCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCA 4024
Db       1568 GGAAGGTCTTTAATTACGATAACGATCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTA 1627

Qy       4025 ATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAG 4084
Db       1628 ATCCCCGCGATGTCGAAGGTTATCGTCAGTTTCTGGACTATTCACGCGCGGTGTTTAAAG 1687

Qy       4085 AGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCG 4144
Db       1688 AAGGCTATCTAAAGCTCGGTACTGTCCCTTTTTATCGTTCAGAGACATGCTTCGCGCCG 1747

Qy       4145 CGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTA 4204
Db       1748 CACCTCAACTGGCGAAACTGCAGGCATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACA 1807

Qy       4205 TTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATC 4264
Db       1808 TCGAAGATGAACATCTGCGCCAGGCGTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATC 1867

Qy       4265 CTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGT 4324
Db       1868 CCTTCGCCACCTCATCCATTTATACGTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCT 1927

Qy       4325 GGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACT 4384
Db       1928 GGTTTCCGCGTGGCGGCACCGGCGCATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATC 1987

Qy       4385 TGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCA 4444
Db       1988 TGGGTGGCGAAGTCGTGTTAAACGCCAGAGTCAGCCATATGGAAACGACAGGAAACAAGA 2047

Qy       4445 TTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATG 4504
Db       2048 TTGAAGCCGTGCATTTAGAGGACGGTCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATG 2107

Qy       4505 CCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTG 4564
Db       2108 CAGATGTGGTTCATACCTATCGCGACCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGT 2167

Qy       4565 CGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGA 4624
Db       2168 CCAACAAACTGCAGACTAAGCGCATGAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGA 2227

Qy       4625 ATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGT 4684
Db       2228 ATCACCATCATGATCAGCTCGCGCATCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGC 2287

Qy       4685 TGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACG 4744
Db       2288 TGATTGACGAAATTTTTAATCATGATGGCCTCGCAGAGGACTTCTCACTTTATCTGCACG 2347

Qy       4745 CGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAG 4804
Db       2348 CGCCCTGTGTCACGGATTCGTCACTGGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGG 2407

Qy       4805 CGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGC 4864
Db       2408 CGCCGGTGCCGCATTTAGGCACCGCGAACCTCGACTGGACGGTTGAGGGGCCAAAACTAC 2467

Qy       4865 GCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAG 4924
Db       2468 GCGACCGTATTTTTGCGTACCTTGAGCAGCATTACATGCCTGGCTTACGGAGTCAGCTGG 2527

Qy       4925 TGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCT 4984
Db       2528 TCACGCACCGGATGTTTACGCCGTTTGATTTTCGCGACCAGCTTAATGCCTATCATGGCT 2587

Qy       4985 CGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCG 5044
Db       2588 CAGCCTTTTCTGTGGAGCCCGTTCTTACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCG 2647

Qy       5045 ATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGC 5104
Db       2648 ATAAAACCATTACTAATCTCTACCTGGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTC 2707

Qy       5105 CCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAAT 5164
```

```
Db        2708 CTGGCGTCATCGGCTCGGCAAAAGCGACAGCAGGTTTGATGCTGGAGGATC---TGATTT 2764

Qy        5165 GAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTT 5224
               ||||   | | | |||||  | || |      |||| |||| || |||||||| |||||
Db        2765 GAATAATCCGTCGTTACTCAATCATGCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTT 2824

Qy        5225 CGCCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGC 5284
                || || ||| | ||| | ||||||||   || || ||||||   |||||||| || ||
Db        2825 TGCGACAGCCTCAAAGTTATTTGATGCAAAAACCCGGCGCAGCGTACTGATGCTCTACGC 2884

Qy        5285 GTGGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCA 5344
                ||||| || || || || |||| ||||| |  || ||||||||||      | |  |||
Db        2885 CTGGTGCCGCCATTGTGACGATGTTATTGACGATCAGACGCTGGGCTTTCAGGCCCGGCA 2944

Qy        5345 GCATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGC 5404
               || |||| |  ||    |  |   |||| ||  ||| ||  | ||  ||||| ||||  ||
Db        2945 GCCTGCCTTACAAACGCCCGAACAACGTCTGATGCAACTTGAGATGAAAACGCGCCAGGC 3004

Qy        5405 CTACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCAT 5464
               |||    ||  |||| ||| | ||||||||||||  ||| ||||||||||||||| ||
Db        3005 CTATGCAGGATCGCAGATGCACGAACCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGC 3064

Qy        5465 TCACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACG 5524
               |||  |  |  | |         ||||||||||||||||||||||||||| |||||| |||
Db        3065 TCATGATATCGCCCCGGCTTACGCGTTTGATCATCTGGAAGGCTTCGCCATGGATGTACG 3124

Qy        5525 CAACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGT 5584
               | | |  || |||       | ||||| |||||||| || ||||||||||| || |||||
Db        3125 CGAAGCGCAATACAGCCAACTGGATGATACGCTGCGCTATTGCTATCACGTTGCAGGCGT 3184

Qy        5585 GGTCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGC 5644
                ||||| |||||||||||||   | ||||||||||| ||  | ||   ||| || | |||
Db        3185 TGTCGGCTTGATGATGGCGCAAATCATGGGCGTGCGGGATAACGCCACGCTGGACCGCGC 3244

Qy        5645 CTGCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGC 5704
               ||| ||  | || ||||| || ||| | || || ||||| ||||| |||| || |||||
Db        3245 CTGTGACCTTGGGCTGGCATTTCAGTTGACCAATATTGCTCGCGATATTGTGGACGATGC 3304

Qy        5705 CGAAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGA 5764
                 |    || ||||| ||||||||||  |  |||||| || || |  ||  |   |   ||
Db        3305 GCATGCGGGCCGCTGTTATCTGCCGGCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGA 3364

Qy        5765 TACGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGC 5824
                |      | |||||  || | |||   ||||||   |    | ||    ||||| |||
Db        3365 GAATTATGCGGCACCTGAAAACCGTCAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCA 3424

Qy        5825 GGAGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTC 5884
               ||| || ||||| || |||   || ||   | ||||  |  | || || || ||||| ||
Db        3425 GGAAGCAGAACCTTACTATTTGTCTGCCACAGCCGGCCTGGCAGGGTTGCCCCTGCGTTC 3484

Qy        5885 GGCGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAATTGGCGTCAAAGTTCAGCA 5944
                || ||||| |||||||||||       ||||| ||  |||| || |||||||| | ||
Db        3485 CGCCTGGGCAATCGCTACGGCGAAGCAGGTTTACCGGAAAATAGGTGTCAAAGTTGAACA 3544

Qy        5945 CGCCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCT 6004
                ||||||  ||| |||||||||   ||||||   || |  |     |||||| |  ||||
Db        3545 GGCCGGTCAGCAAGCCTGGGATCAGCGGCAGTCAACGACCACGCCCGAAAAATTAACGCT 3604

Qy        6005 GCTGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCC 6064
               |||| ||   |   | |||  |||  | ||||| ||  ||      || |   || || ||
Db        3605 GCTGCTGGCCGCCTCTGGTCAGGCCCTTACTTCCCGGATGCGGGCTCATCCTCCCCGCCC 3664

Qy        6065 GGCTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCT 6124
                ||   ||| |||||||| || |      | ||   || || |  |  || ||||| ||
Db        3665 TGCGCATCTCTGGCAGCGCCCGC------TCTAGCGCCATGTCTTTCCCGGAGCGTCGCC 3718

Qy        6125 TGCAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCG 6184
               || || ||   ||| || ||||| |||||||| ||||| || || || || |||||    ||
Db        3719 TGAAGTTTTGACAGGGGCGGCGCATAGAGGAAGCCAAAAGAAACACAACCTTCTTTGCCC 3778

Qy        6185 CGCACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGG 6244
               |  || || |||||||| ||||||||||| || || || || || ||||| ||||| ||
Db        3779 CTGACGGCGTGATGCATACGGTGCGCCATATACAACCGTTTGAGGTAGCCCTTGCGTGGA 3838

Qy        6245 ATATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCG 6304
               ||||||||||| ||||| ||||||||||||||| || |||||||||||| || |||||
Db        3839 ATATAGCGGAATGGCCAACGTTGATGCACCAGCCCGTCGTGCACCATAAAATAGAGTAAT 3898

Qy        6305 CCGTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACA 6364
               || |||| |||||| || || ||||||||||| ||||||||||| |||        || | |
Db        3899 CCATACGCCGTCATACCTGCGCCAATCCACTGGAGCGGCCACATTCCTGTACTGCCCAGA 3958

Qy        6365 TAAATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAAC 6424
               ||||||||||  ||||  | | | |||||  ||||| |||||||||||||| |  ||||||
Db        3959 TAAATCAGCAGGATCGATAATGCAGCAAAAACCACGGCATAAAGATCGTTAACTTCAAAC 4018

Qy        6425 TTACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATG 6484
                 |||  |  |||||||||| || || |||||||| ||||| | ||||| ||||||||||||
Db        4019 GCACCTTTACGCGGTTCATGATGTGAAAGATGCCATCCCCAACCCCAGCCGTGCATGATG 4078

Qy        6485 TATTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGC 6544
               ||||| || | ||| || || |  | ||||||  | | ||||||  | || | ||| ||
Db        4079 TATTTGTGTGCCAGTGCAGCAATCACTTCCATGCCAATCACGGTAACGAAAACGATCAGG 4138

Qy        6545 ACGTTCCATAACCAGAGCATTGTTCGTCCATTTGTGGAAAAGGGAAGTA 6593
                | ||||| | ||| | |||  ||  |||  | | |      || || |
Db        4139 GCATTCCAAATCCACAACATAATTTCTCCGGTAGAGACGTCTGGCAGCA 4187

RESULT 8
BX571873
LOCUS       BX571873              349107 bp    DNA     linear   BCT 26-SEP-2003
DEFINITION  Photorhabdus luminescens subsp. laumondii TTO1 complete genome;
            segment 15/17.
ACCESSION   BX571873 BX470251
VERSION     BX571873.1  GI:36787441
KEYWORDS    complete genome.
SOURCE      Photorhabdus luminescens subsp. laumondii TTO1
  ORGANISM  Photorhabdus luminescens subsp. laumondii TTO1
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1
  AUTHORS   Duchaud,E., Rusniok,C., Frangeul,L., Buchrieser,C., Taourit,S.,
            Bocs,S., Boursaux-Eude,C., Chandler,M., Dassa,E., Derose,R.,
            Derzelle,S., Freyssinet,G., Gaudriault,S., Givaudan,A., Glaser,P.,
            Medigue,C., Lanois,A., Powell,K., Siguier,P., Wingate,V.,
            Zouine,M., Boemare,N., Danchin,A. and Kunst,F.
  TITLE     Complete genome sequence of the entomapathogenic bacterium
            Photorhabdus luminescens
  JOURNAL   Nat. Biotechnol. 11 (1) (2003) In press
REFERENCE   2
  AUTHORS   Duchaud,E., Frangeul,L., Rusniok,C. and Kunst,F.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
            rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
            lfrangeu@pasteur.fr, fkunst@pasteur.fr
FEATURES             Location/Qualifiers
     source          1..349107
                     /organism="Photorhabdus luminescens subsp. laumondii TTO1"
                     /mol_type="genomic DNA"
                     /strain="TTO1"
                     /db_xref="taxon:243265"
     gene            complement(80..3225)
                     /gene="tccC1"
                     /locus_tag="plu4167"
     CDS             complement(80..3211)
                     /gene="tccC1"
                     /locus_tag="plu4167"
                     /function="Unknown, probable insecticidal toxin"
                     /note="Highly similar to insecticidal toxin complex
                     protein TccC of W14."
                     /codon_start=1
                     /transl_table=11
                     /product="Insecticidal toxin complex protein TccC1"
                     /protein_id="CAE16539.1"
                     /db_xref="GI:36787442"
                     /translation="MSTPDTALYTQTPTVSVLDNRGLSIRDIGFHRVVIGGDIDTRVT
                     RHQYDARGYLNHSIDPRLYDAKQADNSVKPNFVWQYDLAGHALRTESVDAGRTVALND
                     IEGRPVMTMNATGVRQTRRYEGNTLPGRLLSVSEQVFDQETSQVTERFIWAENTTAEK
                     EYNLSGLCIRHYDTAGVTRLMSQSLAGAILSQSHQWLAEGQEANWSGDDETVWQRMLA
                     SEVYTTQSTTNAIGALLTQTDAKGNIQRLAYDIAGQLKGSWLTVKGQREQVIVKSLSW
                     SAAGHKLREEHGNGVITEYSYEPETQRLMGITTRRAEESQSEARVLQDLRYEYDPVGN
                     VISIHNDAEATRFWRNQKVEPENRYVYDSLYQLMSATGREMANIGQQSNQLPSPAMPV
                     PADNSTYTNYQRTYSYDRGGNLVQIRHSSPATQNSYTTEMTVSSRSNRTVLSTLTTDP
                     TQVDGLFDAGGHQKTLIPGQNLDWNMRGELQRVTPVSRENSSDSEWYRYSSDGVRLLK
                     VSEQQTGNSTQVQRVTYLPGLELRITGVADKTTQNLQVITVGEAGRAQVRVLHWESGK
                     PADIDNNQVRYSYDNLLGSSQIELDSEGQILSQEEYYPYGGTAVWAAKNQTEASYKFI
                     RYSGKERDATGLYYYGYRYYQPWVGRWLSADPAGTVDGLNVYRMVRNNPITLTDHDGL
                     APSPNRNRNTFWFASFLFRKPDEGMSASMRRGQKIGRAIAGGLAIGGLAAAIAGTAGA
                     ALPVILGVAAVGAGIGALMGYNVGSLLEKGGALLARLVQGKSTLVQSAAGAAAGASSA
                     AAYGARAQGVGVASAAGAVTGAVGSWINNADRGIGGAIGAGSAVGTIDTMLGTSSTLT
                     HEVAAAAGGAAGGMITGTQGSTRAGIHAGIGTYYGSWIGFGLDVASNPAGHLANYAVG
                     YAAGLGAEMAINRVIGGGFLGRLLGRAVSPYAAGLARQLVHFSIARPVFEPIFSVLGG
                     LAGGIGTGLHRVMGRDSWVSRGLSAAGSGIDHLAGMIGNQIRGRVLTTTGIANAIDYG
                     TSAIGAARRVISL"
     RBS             complement(3220. .3225)
                     /gene="tccC1"
                     /locus_tag="plu4167"
     gene            complement(3345. .8051)
                     /gene="tccB1"
                     /locus_tag="plu4168"
     CDS             complement(3345. .8039)
                     /gene="tccB1"
                     /locus_tag="plu4168"
                     /function="Unknown, probable insecticidal toxin"
                     /note="Highly similar to insecticidal toxin complex
                     protein TccB of W14 Putative transmembrane protein."
                     /codon_start=1
                     /transl_table=11
                     /product="Insecticidal toxin complex protein TccB1"
                     /protein_id="CAE16540.1"
                     /db_xref="GI:36787443"
                     /translation="MLSTMEKQLNESQRDALVTGYMNFVASTLKGVDGQPVTVEDLYE
                     YLLIDPEVADEVETSRVAQAIASIQQYMTRLVNGSEPGRQAMELSTANEWRDNDSQYA
                     IWAAGAEVRNYAENYISPITRQEKSHYFSELETTLNQNRLDPDRVQDAVLGYLNEFEA
                     VSNLYVLSGYINQDKFDQAIYYFIGRTTTKPYRYYWRQMDLSKNRQDPAGNPVTPNCW
                     NDWQEITLPLSGDTVLEHTVRPVFYNDRLYVAWVERDPAVQKDADGKNIGKTHAYSIK
                     FGYKRYDDTWTAPNTTTLMTQQADESSETQRSSLLIDESKTTLRKVNLLATTDFSIDP
                     TEETDSNPYGRLMLGVFVRNFTGTPHQNRPAVYGYLYCDSAFNHHILRPLSKDFLFST
                     YRDETNDQNCLQFAVYDKKYVITKVISGKPTEDPENTGWVGKVDDLKLGTTGAYVYID
                     QDGLTLHIQTTTNGDFVNRHIFGYNDIVYDSKSGYGFTWSGNEGFYLDYHDGNYYTFH
                     NAAINYYPAGYGGGSVPNGTWALEQRVNEGWAISPLLDAHHTITVQGSYIAWEGETPA
                     SYYLPIPDGTMLLDWSDKINFAISLNKLESVFTSPDWPTLTTIKNFSKIADNRKFYQE
                     INAETADGRNLFKHYSEQTFQLASGATYPTTYTLSEADFSTDPDKNYLQVCLNVVWDH
                     YDRPSGRKGAYSWVSKWFNVYVELQDSKAPDAIPRLISRYDSKRGLVQYLDFWTSSLP
                     AKTRLNTTFVRTLIEKANLGLDSLLDYTLQADPSLEADLVTDGKSEPMDFDGSNGLYF
                     WELFFHLPFLVATRFANEQQFSPAQKSLHYIFDPAMKNKPHNAPAYWNVRPLVEVHSD
                     LSRHLGDPIDPDTQAYAHPVIYQKAVFIAYVSNLIAQGDMWYRQLTRDGLTQARVYYN
                     LAAELLGPRPDVSLSSTWTPQTLDTLAAGQKTALRDFEHQLANSNTVLPALPGNNISH
                     LKLADNGYFNEPLNVLMLSHWDTLDARLYNLRHNLTVDGKPLSLPLYAAPADPVALLA
                     QRAQSGTLANGVSGAILTVPPYRFSAMLPRAYSAVGTLTSFGQNLLSLLERSERAGQE
                     ELAQQQLLDMSSYAITLQQQALDGLAADRLALLASQATAQQRHDHYYTLYQNNISSAE
                     QLVMDTQTSAQSLISSSTGVQTASGALKVIPNIFGLADGGSRYEGVTEAIAIGLMAAG
                     QATSVVAERLATTENYRRRREEWQIQYQQAQSEVDALQKQLDALAVREKAAQTSLQQA
                     KAQQVQIRTMLTYLTTRFTQATLYQWLSGQLSALYYQAYDAVVSLCLSAQACWQYELG
                     DYATTFIQTGTWNDHYRGLQVGETLQLNLLQMEAAYLVRHERRLNVIRTVSLKSLLGD
                     DGFGKLKTEGKVDFPLSEKLFDNDYPGHYLRQIKTVSVTLPTLIGPYQNVKATLTQTS
                     SSTLLTADINGVKRLNDPTGKEGDATHIVTNLRASQQVALSSGINDAGSFELRLEDER
                     YLSFEGTGAVSKWTLTFPRSVDEHIDDKTLKADEMQAALLANMDDVLVQVHYTACDGG
                     ASFANQVKKTLF"
     RBS             complement(8046. .8051)
                     /gene="tccB1"
                     /locus_tag="plu4168"
     gene            complement(8134. .11034)
                     /gene="tccA1"
                     /locus_tag="plu4169"
     CDS             complement(8134. .11034)
                     /gene="tccA1"
                     /locus_tag="plu4169"
                     /function="Unknown, probable insecticidal toxin"
                     /note="Highly similar to insecticidal toxin complex
                     protein TccA of W14."
                     /codon_start=1
                     /transl_table=11
                     /product="Insecticidal toxin complex protein TccA1"
                     /protein_id="CAE16541.1"
                     /db_xref="GI:36787444"
                     /translation="MNQLASPLISRTEEIHNLPGKLTDLGYTSVFDVVRMPRERFIRE
                     HRADLGRSAEKMYDLAVGYAHQVLHHFRRNSLSEAVQFSLRSPFSVSGPDYANQFLDA
                     NTGWKDKAPSGSPEANDAPVAYLTHIYQLALDQEKSGASAIMNTLVERRPDIGALLIN
                     DKAINEVIPQLQLVNEILSKAIQKKLSLTDLEAVNAKLSTTRYPNNLPYHYGHQQIQT
                     AQSVLGTTLQDITLPQTLDLPQNFWATAKGKLSDATAGALTRLQIMASQLSPEQQKIM
                     VETVGQNFYQLNYGDSLLTANSFSDMTVLTDRTSLTVPQVELMLCSTVGGSTVVKSDN
                     VSSGDTTATPFAYGARFIHAGKPQAITLSRSGAEAHFALTVNNLTDDKLDRINRTVRL
                     QKWLNLPYEDIDLLVTSAMDAETGNTALSMNDNTLRMLGVFKHYQAKYGVTAKQFAGW
                     LRVVTPFAITPATPFLDQVFNSVGTFDTPFVIDNQDFVYTLTTGGDGARVKHISTALG
                     LNHRQFLLLADNIARQQGNITQSTLNCNLFVVSAFYRLANLARTLGTNPESFCALVDR
                     LDAGTGIVWQQLAGKPIIMVPQKDSPLAADILSLLQALSVIAQWQQQRDLPIETLLLL
                     LSSGSTFTAQGTDDQLNFIRQVWQNLGSTFVSATLLSRSGAPLVDTNGHAIDWFALLS
                     AGNSPLIDQVGLVTDAGIESVITMVVKTQSLSDEDKKLAITTLTNTLYQAQQTQQGVA
                     VSLLAQTLNVSQSLPALLLRWSGQTTYQWLSATWALKDTVKIAADIPAGYLSLLREVV
                     RRSLLTQQFTLSPAMVQTLLDYPAYFGASAEVVTDISLWMLYTLSRYSDLLLQVGKAG
                     GGTEDDVLAYLRVANAATPLSQADAAQTLATLLGWETNELQAAWEVLGGIAKTTPQLD
                     TLLRLQQAQSQTGLGVTQQQQGYVLSRDSDYTLWQSTGQAMVAGVSHVKGSH"
     gene            complement(11462. .11943)
                     /locus_tag="plu4170"
     CDS             complement(11462. .11929)
                     /locus_tag="plu4170"
                     /function="Unknown"
                     /note="unnamed protein product; Similar to pyocin S3
                     immunity protein and to Photorhabdus luminescens proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE16542.1"
                     /db_xref="GI:36787445"
                     /translation="MAINIEALINSLGKTYQEIFDEGLIPYKTKPSGYPGDSDISLDM
                     AKEGIFLSFLRENKVLTEITLTFINHKRPNFIFPNKLPSPLIPLMSRKLIHQHFGKPE
                     KSLPPRKRLTKEIGWTELYTLLDFRIPTSMQVDYDLQERVRLVTFLPTSEIRW"
     RBS             complement(11938. .11943)
                     /locus_tag="plu4170"
     gene            complement(12038. .12505)
                     /locus_tag="plu4171"
     CDS             complement(12038. .12505)
                     /locus_tag="plu4171"
                     /function="Unknown"
                     /note="unnamed protein product; Similar to pyocin S3
                     immunity protein and to Photorhabdus luminescens proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE16543.1"
                     /db_xref="GI:36787446"
                     /translation="MTVNVDALINSLGKSYQEIFDEGLIPYKTKPTGYPGASFIALNM
                     AKEGMHLAFKRDGKILFAVELFLLDQKRPLYQFPNELPSPLKPLMTRVWVHEQFGKPE
                     KALPPRKFLKKDVGWTELYTLLDFRIPTSMQVDYDLLEQVKSVAFLSTSEVLW"
     gene            complement(12597. .13081)
                     /locus_tag="plu4172"
     CDS             complement(12597. .13064)
                     /locus_tag="plu4172"
                     /function="Unknown"
                     /note="unnamed protein product; Similar to pyocin S3
                     immunity protein and to Photorhabdus luminescens proteins"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAE16544.1"
                     /db_xref="GI:36787447"
                     /translation="MIDVIEIINSLGKTADMLVENKLIPEEKFEFAFEGYSRFSVEPE
                     TGLTLVFDIISTVLISVQFTLINTVDDTGTYSGDMPAPFLHKMNKAIVRAFMGEPDYS
                     SGPERIPVIGLVGGFDAYTRKLNDQYPNTEIRFLYQTDLRVDALLFEQSVNND"
     RBS             complement(13076. .13081)
                     /locus_tag="plu4172"
     gene            complement(13174. .13652)
                     /locus_tag="plu4173"
```

```
CDS             complement(13174..13635)
                /locus_tag="plu4173"
                /function="Unknown"
                /note="unnamed protein product; Some similarities with
                pyocin S3 immunity protein and similar to Photorhabdus
                luminescens proteins"
                /codon_start=1
                /transl_table=11
                /protein_id="CAE16545.1"

  Query Match             19.6%;  Score 1374.8;  DB 1;  Length 349107;
  Best Local Similarity   62.0%;  Pred. No. 1.3e-278;
  Matches 2211;  Conservative    0;  Mismatches 1347;  Indels    8;  Gaps    2;

Qy       2532 AATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGC 2591
              | |  ||| ||||| |||| || |  ||| |||| || ||  |||| ||  ||||| | |
Db     202944 ACTGGGATCTGATTCTGGTGGGAGGAGGATTGGCCAACGGACTGATCGCAATGCGTTTCC 203003

Qy       2592 GTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCA 2651
                || |   | ||||| |||  |    || |  | || | |   |        | |||||
Db     203004 AGCAGTGTAAACCACATCTGCGAGTGTTGCTTATTGAAAACACAGAAACAATAGGAGGCA 203063

Qy       2652 ATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAAC 2711
              |||| || ||||| ||||||||    |||||| |  |  |   ||| |   ||| |   ||
Db     203064 ATCACACGTGGTCATTTCATCAACATGATCTTACTGAGGCGGAACATGAGTGGATAGCAC 203123

Qy       2712 CGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATC 2771
              |||||||||||   || |||||||||||  | ||   |||||| ||  | |  ||||
Db     203124 CGCTGATTACCTATCGCTGGTCAGGTTACGACGTCATTTTTCCAGCATTTCAACGCACAT 203183

Qy       2772 TGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGA 2831
              ||          ||||     ||| ||||     | |||||  ||   || | ||
Db     203184 TGCCACATTCATATTTCAGTATCACATCCCAACACTTTGCAAGCATACTCCATGCATATT 203243

Qy       2832 TGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGC 2891
              |||| ||   | |   |||      |   |||||  || | |  ||   | | ||
Db     203244 TGGGCGAACGTATACAGACTCGTTTATTGGTACAGGAGCTGACTCCACAGAAAGTTTACT 203303

Qy       2892 TGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGA 2951
              |   ||| |||      ||   |||    |  || |||||||||  | |||  ||  ||
Db     203304 TACAGGACGGCTCGTCTCTAAGTGCTGGCGCAGTCATTGATGGGAGAGGCTGGCGACCAG 203363

Qy       2952 CGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGC 3011
                |||  | |   |   ||    ||||  ||| |||| || || ||| ||||||   ||
Db     203364 GACCATTTATAGGGAGTGGCACCCAGGCATTTTTTGGTCAGGAATGGGAGCTGGAAGAGT 203423

Qy       3012 CGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATC 3071
              |||||    |    || |||||  | |||||| | |  || |  ||| |  | |||||||
Db     203424 CGCACTCTTTAACCCACCCGATTTTAATGGATACTAGTGTGGGACAGGATACAGGTTATC 203483

Qy       3072 GTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACG 3131
              | ||| ||||    |||||| ||  | |   ||| ||  ||||| || || || |||||||
Db     203484 GATTTATCTATGTCCTGCCGTTCTCCTCAACTCGTCTGCTGATAGAGGACACTCATTACG 203543

Qy       3132 TTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATC 3191
              || | | |    | |  |   ||       | |        ||||| ||  || || ||
Db     203544 TTGATCGGGGGCCACCTGATAAGGCTTTGTCGCAGGCTACTATCGCAGAGTACGCGAAGA 203603

Qy       3192 AGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCC 3251
              | || || ||||  ||| ||||  ||  | || ||||||  |||     | || ||||| |
Db     203604 AACATGGATGGAAACTGGGTAAACTCATTCGAGAAGAGAGCGGTTGTCTTCCAATTACAC 203663

Qy       3252 TGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGC 3311
              | |  ||  |  |       ||||||  |||||     || ||  |   | | ||  | |
Db     203664 TTACGGGAGATTTTACCTCTTTCTGGGCACAGCTAGCAGGACAGCCCACCTGTGGGTTAC 203723

Qy       3312 GCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGG 3371
              | || |   |||||||  ||| || || |||||  ||||   |||  |   | | || |
Db     203724 GTGCAGCTTTGTTTCACCCCACAACAGGCTACTCTCTGCCACACGCTATTCGGTTGGCAG 203783

Qy       3372 AGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCT 3431
              |     | |  || |||  |         |    | |  |   |       ||      |
Db     203784 ATCGTATTGTTGCTCTGCCGGAGCTTACCGATACCTCCTTATTCATTACCCTCAGGGATT 203843
```

```
Qy       3432 TTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTT 3491
                || || || |||||||   | ||||| |||||||||| ||||| ||||||||| ||
Db     203844 ACGCACGACAACAGTGGCAACACCAGCGCTTTTTCCGTCTTCTAAATCGCATGCTCTTCC 203903

Qy       3492 TGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCG 3551
              | || ||  | || || || || ||||  || |||||||||||| || |  ||     |
Db     203904 TCGCTGGGGATCCACAACAACGTTGGCAGGTAATGCAACGTTTCTATCAACTTTCCCCAA 203963

Qy       3552 GGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGT 3611
                 | |||   || ||||| || | ||||||         ||| ||  | ||||||||
Db     203964 ATCTGATTGCACGTTTTTATGCGGAGCAACTTAATTCCGTCGACAAGGCCCGGATTCTCA 204023

Qy       3612 GCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAA-----TTCTGT 3666
                || || || || ||||| ||    |  || ||    ||  |||| ||         |
Db     204024 TAGGTAAACCACCAGTGCCGATAAAGGGTGCTCTAAAGGCAATGTTTAAACAACACAAGA 204083

Qy       3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
               |   |||||     |  ||||      |    || |||||| || || ||||| || |||
Db     204084 AGCTTCAGGGTTTTTATTATGATTAAAGCGCTGGTAATTGGTGCTGGTTTTGGTGGGCTG 204143

Qy       3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
              || ||||| ||  | || ||  | ||||| || || ||      |  | ||||| || ||
Db     204144 GCACTGGCAATAAGGCTCCAGTCTGCGGGGATTCCGACATGTATTTTGGAGCAACGGGAT 204203

Qy       3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
              ||||| || |||||||| ||||| | | ||||    || || ||||| |||||||| |||
Db     204204 AAACCAGGTGGACGCGCTTATGTTTATAAGGAACAGGGATTCACCTTTGATGCCGGCCCC 204263

Qy       3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
              || || ||||||| ||| |  | ||| |||||| ||||| || | ||| || |||    |
Db     204264 ACCGTAATCACCGCTCCTAATGTCATTGAAGAATTGTTTACCCAGGCTGGTAAACGTATG 204323

Qy       3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
                 |||||||||||  |  |  || |||   || || |||| || || || |||    ||
Db     204324 GCTGATTACGTCGATTTACTTCCCGTACACCCTTTTTATCGGCTCTGTTGGGAGTCGGGT 204383

Qy       3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
              ||   | ||||||||||||||| ||||||  |  |||||    || |||    |||||||||
Db     204384 AAGATGTTTGATTACGACAATGATCAGCAACATCTGGAAGCACAAATCCATACGTTCAAT 204443

Qy       4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
              ||   ||||||  | || ||||| |  ||||| | ||||||| || ||||  ||||  ||
Db     204444 CCAAGAGATGTTAACGGGTATCGACGCTTTCTAGACTATTCCCGGGAAGCCTTTAATGAA 204503

Qy       4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
              || || |||||||| || |||||||| ||| ||     | || |||||| || | |  ||
Db     204504 GGCTACCTGAAACTTGGAACGGTGCCCTTCCTCTCTTTTCGCGACATGTTGAGTGCGGCT 204563

Qy       4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
              || ||||||  |||||| || |||||||| || || || ||   ||||||    |||||
Db     204564 CCACAGTTGATACGTCTACACGCATGGCGTAGTGTTTATAGTCAGGTGGCACGCTTTATC 204623

Qy       4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
               |||| ||    ||||| ||||| ||||| || |||||  |  || | || ||||| ||
Db     204624 AAGGATGAGAGCCTGCGCCAGGCATTTTCATTTCACTCGCTATTGATAGGAGGTAACCCC 204683

Qy       4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
              |||||  | || || |||||||| || || || || |||||| || |||||||| || |||
Db     204684 TTTGCTGCATCTTCTATCTATACTTTGATCCACGCACTGGAACGAGAATGGGGAGTCTGG 204743

Qy       4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
              |||||||| || || ||| ||||| |||||  |||  |||| |  | ||||    ||| |
Db     204744 TTTCCGCGAGGAGGAACCAGCGCTCTGGTTGAGGCAATGGTGAAATTGTTTACTGACATT 204803

Qy       4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
              || || ||  |    || |||||  | || |  |   |     |||   ||||||||  |
Db     204804 GGTGGAGAAATTGAGCTTAATGCAAAGGTAAAACGCTTCACCACCCATGGCAATCGGGTC 204863

Qy       4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
              |  || || || ||||  |  ||||||    |       | || |||||||||| ||||||
Db     204864 ACAGGTGTCCAATTAGCTGATGGACGAAATATGACATGTGACGTTGTGGCCTCAAATGCC 204923
```

```
Qy       4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
              || ||  | ||||||||  |  | ||| |  | || |||||| | | |||    || ||
Db     204924 GATGTTATTCATACCTATAAGCACCTGTTAGGTCAGCATCCGGTCGGAATAGCCCGCGCT 204983

Qy       4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
                | | |||  |||||| || ||||||||||| ||||| || || ||||| || |||||
Db     204984 CGAACACTGGTGCGTAAACGAATGAGCAACTCACTGTTCGTGCTTTATTTCGGTCTGAAC 205043

Qy       4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
              || |  |||   ||  | || || || || || ||||| || || |||   ||  |
Db     205044 CATCACCATACGCAATTAGCCCATCATACAGTTTGCTTTGGTCCACGCTATAAGGAACTC 205103

Qy       4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
              ||||| || |||||  |      ||  ||| |||| ||||||||||| || || || ||
Db     205104 ATCGAAGATATTTTTTATCATGACCGACTGTCAGAGGATTTTTCACTCTATCTCCATGCC 205163

Qy       4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
              ||||     |  || || || || || || |  || ||||  || |  || ||  | ||
Db     205164 CCCTCTGTTACTGACCCATCCCTTGCTCCCCGAGGATGCGCAAGTTACTACGTTCTGGCC 205223

Qy       4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
              || |  || ||| | || || ||| |  |  ||||| |    || || || ||||| ||
Db     205224 CCTGTACCACATTTGGGGACAGCTAATTTAAACTGGGATATAGAGGGGCCACGCTTACGT 205283

Qy       4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
              || || ||||| || ||||||||  || |||| ||||| ||  | |   |||| |||||
Db     205284 GACCGTATTTTCGCGTATCTGGAAAAGTACTATATGCCCGGTCTGCTCAAGCAGTTAGTC 205343

Qy       4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
                 ||  | || ||||| || ||||||||||| |||   ||  ||||||||| ||| |||
Db     205344 GTTCATCGTATTTTTACCCCATTTGATTTTCGTGACCAACTCAATGCCCATCTCGGTTCG 205403

Qy       4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
              || || ||  | || ||| |  |||| |||||||| ||||||||| || || || || ||
Db     205404 GCCTTCTCCTTTGAACCGTTGCTGACACAAAGCGCTTGGTTCCGACCACACAATCGGGAC 205463

Qy       5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
                |   ||    ||||| ||||| || || || || |||||||| || || ||  | |||
Db     205464 AACCGCATTGATAATCTGTATCTCGTCGGCGCAGGCACGCATCCTGGAGCAGGTATTCCC 205523

Qy       5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
              |||||||| || || || || ||||||||||   ||||| | ||||||  |||  |||||
Db     205524 GGCGTGATTGGATCAGCAAAAGCCACCGCCACATTGATGTTAGAGGATATCGCTAAATGA 205583

Qy       5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
              |||    ||   ||||| | || || || || |  |||| |   ||||| || ||||| |
Db     205584 ATCCTGCGC---TACTTAAACACGTTACTCAGATAATGGAGCAAGGCTCAAAAAGTTTTG 205640

Qy       5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
              |||  | | ||   || || ||| || | || || | ||||||  |||||||||| || |
Db     205641 CCAGTGTCACCCGACTTTTCGATACAGCAACACGACACAGCACAATGATGCTGTACGCCT 205700

Qy       5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
              |||| ||| | || |||||| | || || ||||||  ||||||   |      |
Db     205701 GGTGCCGTTATTGTGATGATATAATAGACGGGCAAGAGCTGGGGAGACAAATAAGCAGTG 205760

Qy       5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
               ||      |   ||| |  |      | ||    ||||||      ||    |  || |
Db     205761 TTGATAAGTATAGCGCCCGCGAGAAACTTCAAATGCTGCAATACCTGACAAAGCAGGCTT 205820

Qy       5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
              ||  |||   ||  |||   ||||| || |||   || ||||||  ||| ||  | | |
Db     205821 ACGACGGTTTGCCGATGACCGAACCTGCTTTTGCAGCCTTTCAGACAGTAGCTCTGAGTA 205880

Qy       5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
              |  |  | || || |||| ||| || || ||||||||||| ||||| ||||||| || |
Db     205881 ATGAAATCCCTCAGCAACAGGCCTTCGAACATCTGGAAGGATTCGCAATGGATGTACTCT 205940

Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
                || | ||||   |  || ||||| || ||| | || ||||| || |||||||| || |
Db     205941 GTGAGCCTTACAGAACGTTGGATGATACTCTGAGATATTGCTACCATGTCGCGGGAGTAG 206000

Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
              | ||  ||||||||||  | || ||||| || ||||| |     ||||| ||||  ||||
Db     206001 TTGGGCTGATGATGGCAAGGGTTATGGGGGTTCGCGAGGCGAGTGTGCTGGATCGGGCCT 206060

Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
              ||||| ||||| | || || ||| | ||||||||||| || || ||| | || |||||
Db     206061 GCGATCTAGGAATTGCTTTTCAGTTAACTAACATTGCTCGGGATATTATTGAGGATGCGA 206120

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
              |    || || |||||||| |||| |   |||||  | || |  ||  ||    | || |
Db     206121 AGGCCGGGCGTTGCTATCTTCCGCTAGAATGGCTTCACCAAGAAGGGCTAATGCCAGACA 206180

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
              | || | |    |  || | ||| | ||||| ||     ||||| | || ||  | | ||
Db     206181 CACTTATTTATACCGAAAACCGTCCTGCGCTAGCTCGGGTGGCATCCCGATTGATTGTGG 206240

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
              | |||||| | ||||||  ||| || |   | ||  |    ||||||||  |||| || |
Db     206241 AAGCGGAATCTTATTATACCTCCGCTCTGACAGGGCTTGTTGGTTTACCTTTGCGTTCAG 206300

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | |||| ||||||| | ||||  ||  | ||||| ||||| ||| | ||||| || || |
Db     206301 CCTGGGTCATCGCTTCTGCTCATGGTATCTATCGTGAAATCGGCATTAAAGTGCAACAAG 206360

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              | || || |  |||||||| |   | ||  | || | |   || || ||  |  |  |
Db     206361 CAGGAGTACGGGCCTGGGAATGCAGACAAAGAACTAATCGTGGAGAGAAGGTTACCTTAT 206420

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              |  ||   || |||     |||| | |  || || ||| |     |    || ||
Db     206421 TAATGGCTGGCGCAATGAAGGCGCTTATATCTCGGGTGGCGATAACAACGCCTCGCGATC 206480

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGA 6092
              |     |||||||||| ||||| || |
Db     206481 CCAAATTGTGGCAGCGACCTCATTAA 206506
```

```
RESULT 9
ERWCRTIB
LOCUS       ERWCRTIB                2415 bp    DNA     linear   BCT 11-APR-2001
DEFINITION  Pantoea agglomerans phytoene dehydrogenase (crtI) and prephytoene
            pyrophosphate synthase (crtB) genes, complete cds.
ACCESSION   M38423
VERSION     M38423.1  GI:148401
KEYWORDS    .
SOURCE      Pantoea agglomerans
  ORGANISM  Pantoea agglomerans
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Pantoea.
REFERENCE   1  (bases 1 to 2415)
  AUTHORS   Armstrong,G.A., Alberti,M. and Hearst,J.E.
  TITLE     Conserved enzymes mediate the early reactions of carotenoid
            biosynthesis in nonphotosynthetic and photosynthetic prokaryotes
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 87, 9975-9979 (1992)
  MEDLINE   91088634
COMMENT     Draft entry and computer-readable sequence for [Proc. Natl. Acad.
            Sci. U.S.A. 87, 9975-9979 (1990)] kindly submitted by M.Alberti,
            11-SEP-1990.
FEATURES             Location/Qualifiers
     source          1..2415
                     /organism="Pantoea agglomerans"
                     /mol_type="genomic DNA"
                     /strain="Eho10"
                     /db_xref="taxon:549"
     gene            11..1489
                     /gene="crtI"
     CDS             11..1489
                     /gene="crtI"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene dehydrogenase"
                     /protein_id="AAA24820.1"
                     /db_xref="GI:148402"
                     /translation="MKKTVVIGAGFGGLALAIRLQAAGIPTVLLEQRDKPGGRAYVWH
                     DQGFTFDAGPTVITDPTALEALFTLAGRRMEDYVRLLPVKPFYRLCWESGKTLDYAND
                     SAELEAQITQFNPRDVEGYRRFLAYSQAVFQEGYLRLGSVPFLSFRDMLRAGPQLLKL
                     QAWQSVYQSVSRFIEDEHLRQAFSFHSLLVGGNPFTTSSIYTLIHALEREWGVWFPEG
                     GTGALVNGMVKLFTDLGGEIELNARVEELVVADNRVSQVRLADGRIFDTDAVASNADV
                     VNTYKKLLGHHPVGQKRAAALERKSMSNSLFVLYFGLNQPHSQLAHHTICFGPRYREL
                     IDEIFTGSALADDFSLYLHSPCVTDPSLAPPGCASFYVLAPVPHLGNAPLDWAQEGPK
                     LRDRIFDYLEERYMPGLRSQLVTQRIFTPADFHDTLDAHLGSAFSIEPLLTQSAWFRP
                     HNRDSDIANLYLVGAGTHPGAGIPGVVASAKATASLMIEDLQ"
     gene            1486. .2415
                     /gene="crtB"
     CDS             1486. .2415
                     /gene="crtB"
                     /note="putative"
                     /codon_start=1
                     /transl_table=11
                     /product="prephytoene pyrophosphate synthase"
                     /protein_id="AAA24821.1"
                     /db_xref="GI:148403"
                     /translation="MSQPPLLDHATQTMANGSKSFATAAKLFDPATRRSVLMLYTWCR
                     HCDDVIDDQTHGFASEAAAEEEATQRLARLRTLTLAAFEGAEMQDPAFAAFQEVALTH
                     GITPRMALDHLDGFAMDVAQTRYVTFEDTLRYCYHVAGVVGLMMARVMGVRDERVLDR
                     ACDLGLAFQLTNIARDIIDDAAIDRCYLPAEWLQDAGLTPENYAARENRAALARVAER
                     LIDAAEPYYISSQAGLHDLPPRCAWAIATARSVYREIGIKVKAAGGSAWDRRQHTSKG
                     EKIAMLMAAPGQVIRAKTTRVTPRPAGLWQRPV"
ORIGIN

  Query Match             17.0%;  Score 1189;  DB 1;  Length 2415;
  Best Local Similarity   68.6%;  Pred. No. 1.6e-239;
  Matches 1655;  Conservative    0;  Mismatches  755;  Indels    3;  Gaps    1;

Qy       3678 AGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGAT 3737
              | |||  ||||||   ||   |||||||||||||||||||||| ||||||||||||||||||
Db          4 ATAAAGGATGAAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGAT 63

Qy       3738 TCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGG 3797
              ||||||||| ||||| || || ||||||    | || |||||||| ||||| || |||||
Db         64 TCGCCTGCAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGG 123

Qy       3798 ACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCAC 3857
               || ||||| || |   | |||   |||||||||||| || ||||| || ||||||||||
Db        124 TCGGGCCTACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCAC 183

Qy       3858 CGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGT 3917
              |||||| | |||  | || | | ||||||| ||||| || |      |    ||||||||
Db        184 CGATCCTACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGT 243

Qy       3918 CGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGA 3977
              |  |||| |||||||||  |||||||| || || ||||||||    || ||    || ||
Db        244 CAGGCTGCTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGA 303

Qy       3978 TTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGT 4037
               || |  ||  |     |   ||| |||  ||||||  ||  |||||| || |  || ||
Db        304 CTATGCTAACGACAGCGCCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGT 363

Qy       4038 AGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAA 4097
               || ||||| || |  |||||| || || || |  |  |||||    ||||| ||| ||
Db        364 CGAGGGCTACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCG 423

Qy       4098 ACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGG 4157
               |||||||  |||||||| ||     | || ||||||||||||| || |||||| ||
Db        424 CCTCGGCAGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCT 483

Qy       4158 ACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCA 4217
                  || || || ||||  |||||||||     |||  ||   ||||||| |||| || ||
Db        484 TAAGCTCCAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCA 543

Qy       4218 TCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTC 4277
              |||||| |||||| || || ||||||||  ||||||| |||||| || || ||  | || ||
Db        544 TCTGCGGCAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTC 603

Qy       4278 ATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGG 4337
               || ||||| ||| | || || || || || ||| || ||||| || |||||| ||    ||
```

```
Db        604 GTCCATCTACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGG 663

Qy       4338 CGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCT 4397
              ||||||||| ||||||||| | ||||||| |   |||||    ||  ||||||| ||| |
Db        664 CGGCACCGGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGAT 723

Qy       4398 GTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCA 4457
                 ||| || |||   ||      ||||||       ||  || ||| | |||   || |
Db        724 CGAACTCAACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCG 783

Qy       4458 GTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCA 4517
              | | | ||  || ||   ||| ||  ||| ||| || |||| || || || |||||||| |
Db        784 GCTGGCGGATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAA 843

Qy       4518 TACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAA 4577
               |||||  | || |||||  ||||||||||| |||    ||| || ||| || ||||| |
Db        844 CACCTATAAAAAGCTGCTCGGCCACCATCCGGTGGGGCAGAAGCGGGCGGCAGCGCTGGA 903

Qy       4578 GCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGA 4637
              ||| ||| |||||||||||||||||||||| ||||| || |||||||| ||||| |||
Db        904 GCGCAAGAGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTC 963

Qy       4638 ACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGAT 4697
               ||||| |||||||| ||| ||||||||||| || || || || ||| |||||||| |||||
Db        964 CCAGCTGGCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGAT 1023

Qy       4698 TTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAG 4757
               || | | ||||  ||||||| || || || || || ||||||||| ||||||||   |
Db       1024 CTTTACCGGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGAC 1083

Qy       4758 CGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCA 4817
              |||||| ||||| || || |||||||||| |||||| || ||| | || ||||||||||||
Db       1084 CGATCCCTCGCTCGCGCCTCCCGGCTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCA 1143

Qy       4818 TCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTT 4877
              ||| |||| |||     | ||||||   ||||| || |||    ||||||| || || ||
Db       1144 TCTTGGCAACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTT 1203

Qy       4878 TGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAAT 4937
              ||  || || ||  ||| ||| ||||| ||  | |||   ||  | |||||| ||  | ||
Db       1204 TGACTACCTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGAT 1263

Qy       4938 GTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCT 4997
               ||||| |||   || || | ||||||||| | |||| ||||  || |||| | || ||  |
Db       1264 CTTTACCCCGGCAGACTTCCACGACACGCTGGATGCGCATCTGGGATCGGCCTTCTCCAT 1323

Qy       4998 GGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAG 5057
               |||||| |  |||| ||||||||||||||||||||||| ||||||||   ||| ||
Db       1324 CGAGCCGCTGCTGACCCAAAGCGCCTGGTTCCGCCCGCACAACCGCGACAGCGACATTGC 1383

Qy       5058 CAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGG 5117
              ||| ||||| |||||||| || ||||| || || || |||||| | || ||||  | |
Db       1384 CAACCTCTACCTGGTGGGCGCAGGTACTCACCCTGGGGCGGGCATTCCTGGCGTAGTGGC 1443

Qy       5118 TTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCT 5177
               ||||| || ||||||||||| |||||| | |||||||      |||||  | || |||
Db       1444 CTCGGCGAAAGCCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCG 1500

Qy       5178 TTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCC 5237
               | ||||| || |  ||||| ||||||||    ||||||| | ||||| |||||||| ||
Db       1501 CTGCTTGACCACGCCACGCAGACCATGGCCAACGGCTCGAAAAGTTTTGCCACCGCTGCG 1560

Qy       5238 AAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCAC 5297
              |||||||| ||  |  | || ||||| |||  ||||||||| ||  | ||||| || |||
Db       1561 AAGCTGTTCGACCCGGCCACCCGCCGTAGCGTGCTGATGCTCTACACCTGGTGCCGCCAC 1620

Qy       5298 TGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAA 5357
              |||||||| || ||||| |  || || |  |||   | | ||  |       || |  ||
Db       1621 TGCGATGACGTCATTGACGACCAGACCCACGGCTTCGCCAGCGAGGCCGCGGCGGAGGAG 1680

Qy       5358 GACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCG 5417
              || ||       ||  ||   |  ||||  |     ||||     || |     || ||
Db       1681 GAGGCCACCCAGCGCCTGGCCCGGCTGCGCACGCTGACCCTGGCGGCGTTTGAAGGGGCC 1740
```

```
Qy         5418 CACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCG 5477
                 | ||| | || ||||| ||    || |||||||| |||||| | |  |||    |  ||
Db         1741 GAGATGCAGGATCCGGCCTTCGCTGCCTTTCAGGAGGTGGCGCTGACCCACGGTATTACG 1800

Qy         5478 CAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTAC 5537
                |  |   ||||| | ||||| || || ||||| || ||||| |      |    |  ||
Db         1801 CCCCGCATGGCGCTCGATCACCTCGACGGCTTTGCGATGGACGTGGCTCAGACCCGCTAT 1860

Qy         5538 GCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATG 5597
                |  | ||| || || |||||||| |||||||||||||| ||||||||||| ||| |||||
Db         1861 GTCACCTTTGAGGATACGCTGCGCTACTGCTATCACGTGGCGGGCGTGGTGGGTCTGATG 1920

Qy         5598 ATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGA 5657
                |||||  | || ||||||||||| || ||   |||||| |||| |||||||||| | ||
Db         1921 ATGGCCAGGGTGATGGGCGTGCGGGATGAGCGGGTGCTGGATCGCGCCTGCGATCTGGGG 1980

Qy         5658 CTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGC 5717
                ||||| |||||||| || || || || || || ||| | || ||||| |  | ||  |||
Db         1981 CTGGCCTTCCAGCTGACGAATATCGCCCGGGATATTATTGACGATGCGGCTATTGACCGC 2040

Qy         5718 TGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCA 5777
                |||||||||||       |||||  |  | || ||  |   | | || |      | ||
Db         2041 TGCTATCTGCCCGCCGAGTGGCTGCAGGATGCCGGGCTGACCCCGGAGAACTATGCCGCG 2100

Qy         5778 CCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCC 5837
                | | |  |||| || ||||| ||     |||| | |||  |  | |  |  || || ||
Db         2101 CGGGAGAATCGGGCCGCGCTGGCGCGGGTGGCGGAGCGGCTTATTGATGCCGCAGAGCCG 2160

Qy         5838 TATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATC 5897
                || ||   |||  | |   ||||  |||  | | | |||| |||||  || ||||| |||
Db         2161 TACTACATCTCCTCCCAGGCCGGGCTACACGATCTGCCGCCGCGCTGCGCCTGGGCGATC 2220

Qy         5898 GCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCAC 5957
                || || || ||| |||| || || || || ||  | || ||  |    || || |    |
Db         2221 GCCACCGCCCGCAGCGTCTACCGGGAGATCGGTATTAAGGTAAAAGCGGCGGGAGGCAGC 2280

Qy         5958 GCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGG 6017
                |||||||||   || |||| |||||| |||||||||||| | ||  ||||| ||   |
Db         2281 GCCTGGGATCGCCGCCAGCACACCAGCAAAGGTGAAAAAATTGCCATGCTGATGGCGGCA 2340

Qy         6018 GCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGG 6077
                 | ||   ||  ||     ||     | |  |       ||||||||||| ||||| |||
Db         2341 CCGGGGCAGGTTATTCGGGCGAAGACGACGAGGGTGACGCCGCGTCCGGCCGGTCTTTGG 2400

Qy         6078 CAGCGTCCTCGTT 6090
                ||||||||   ||
Db         2401 CAGCGTCCCGTTT 2413

RESULT 10
AF408848
LOCUS       AF408848                3801 bp    DNA     linear   BCT 29-OCT-2002
DEFINITION  Xanthobacter sp. Py2 zeaxanthin biosynthetic gene cluster, complete
            sequence.
ACCESSION   AF408848
VERSION     AF408848.1  GI:15553718
KEYWORDS    .
SOURCE      Xanthobacter sp. Py2
  ORGANISM  Xanthobacter sp. Py2
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Hyphomicrobiaceae; Xanthobacter.
REFERENCE   1  (bases 1 to 3801)
  AUTHORS   Larsen,R.A., Wilson,M.M., Guss,A.M. and Metcalf,W.W.
  TITLE     Genetic analysis of pigment biosynthesis in Xanthobacter
            autotrophicus Py2 using a new, highly efficient transposon
            mutagenesis system that is functional in a wide variety of bacteria
  JOURNAL   Arch. Microbiol. 178 (3), 193-201 (2002)
  MEDLINE   22176293
   PUBMED   12189420
REFERENCE   2  (bases 1 to 3801)
  AUTHORS   Larsen,R.A. and Metcalf,W.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-2001) Department of Microbiology, University of
            Illinois, 601 S. Goodwin Ave, Urbana, IL 61801, USA
FEATURES             Location/Qualifiers
     source          1. .3801
                     /organism="Xanthobacter sp. Py2"
                     /mol_type="genomic DNA"
                     /strain="Py2"
                     /db_xref="taxon:78245"
     gene            170. .1327
                     /gene="crtY"
     CDS             170. .1327
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /product="lycopene cyclase"
                     /protein_id="AAL01999.1"
                     /db_xref="GI:15553719"
                     /translation="MDIVFVGAGLANCLMAARLAAQRPGLHMLLLEAGESVGGNHSWS
                     CHDSDLTAAQRAFLAPFQSYLWAGHGVHFPAFSRTLKGGYATISSERMAEVMNERLCA
                     AIRTNARVAHVAPDHVVLEGGERIDARAVVDGRGPLASRHLDLGYQTFLGQELRMSRP
                     HGLTRPIIMDARVEQLGGYRFVYVLPLDDDTLLVEDTYYADGPDLPADALRGRISAYA
                     AAQGWAVDYVVREEDGILPIALGGDINAFLAETPSGVAPAGLRAGLFHPTTGYSLPDA
                     MALADSVSALADLSGPALSAAVRSHAAAAWNGRGFFRLLNRMLFRAADPERRYAILQR
                     FYGLSEDLIARFYADRLTLADKARILSGRPPVSVFRALSCLVETKAAPGSP"
     gene            1324. .2838
                     /gene="crtI"
     CDS             1324. .2838
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene dehydrogenase"
                     /protein_id="AAL02000.1"
                     /db_xref="GI:15553720"
                     /translation="MMLDATKTIRRAAVIGSGFGGLSLAIRLQAAGIRTTVFEQRDKP
                     GGRAYVYEQDGFTFDGGPTVITDPTCLEEVFAAAGRKLSDYVELMPVSPFYRLLWPDG
                     RQFDYVNDQAALDAQIATFDKADVEGYRRFLAYSQAVFEEGYLKLGAVPFLEFSSMMK
                     AAPKLVRLEAWRSVYAMVSRFIRDDHLRQAFSFHSLLVGGNPFSTSSIYALIHALERK
                     WGVFFPRGGTGALVRGMVKLFTDLGGEIRLTSPVDEIVVEGQRATAVKLKSGETLPFD
                     LVASNADVVHTYRHLLRGAARGRSEGARLAKTRHSMSLFVTYFGARRTWDHLQHHTVL
                     FGPRYRGLVDEIFKGPNLPDDFSLYLHAPTVTDKSLAPEGCTAFYVLSPVPHLGKADI
                     DWEVEGPLYRDRILAHLEERLLPGLRDSLVTSRILTPFGFRDELSAHQGSAFSVEPLL
                     TQSAWFRPHNRDAKIANLYFAGAGTHPGAGVPGVVGSAKATAGLILADLGVTPR"
     gene            2835. .3770
                     /gene="crtB"
     CDS             2835. .3770
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene synthase"
                     /protein_id="AAL02001.1"
                     /db_xref="GI:15553721"
                     /translation="MSTPAPTVLGASEEAINKGSKSFAAAARLFGPRMREDAVMLYAW
                     CRHCDDVVDGQELGHGRVSSAASPAERLDGLYEETRNAYRGAPSAHPAFAAFAEVVKR
                     NDIPERYPLQLLEGFRMDVEGRRYDTLDDTLTYCYHVAGVVGVMMALIMGARDEVVLD
                     RASDLGLGFQLTNIARDVMEDAAIGRIYLPAAFLAEAGVPEAAIAAPEHRAAVAGVVA
                     RLLDVAEPYYDQALIGMAALPFRAAAAVGAARGVYRAIGIEVRKRGPKAWDTRVSTST
                     AQKAGFLAGGLAQALATRFRGAPPPRPAHLWTRPA"
ORIGIN

  Query Match             16.9%;  Score 1179.8;  DB 1;  Length 3801;
  Best Local Similarity   59.2%;  Pred. No. 1.4e-237;
  Matches 2131;  Conservative    0;  Mismatches 1422;  Indels   48;  Gaps    5;

Qy         2537 GATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAA 2596
                ||  |  | || ||||| || ||||| || ||| || | || |||  ||| ||
Db          173 GACATCGTCTTCGTCGGCGCCGGACTTGCCAATTGCCTTATGGCGGCGCGCCTCGCGGCC 232

Qy         2597 TTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCAT 2656
                  ||  ||    ||  |    ||| |||||||   ||  | |     ||  ||||||||
Db          233 CAGCGTCCCGGCCTTCACATGCTGCTGCTGGAAGCCGGCGAGAGCGTGGGCGGCAATCAC 292

Qy         2657 ACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTG 2716
                | |||||| |  ||  ||||||||||||  ||||  || |    ||  ||    ||  |
Db        293 AGCTGGTCCTGCCACGACAGCGATCTCACTGCCGCGCAGCGCGCCTTCCTCGCTCCCTTC 352

Qy       2717 ATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGAC 2776
Db        353 CAGAGCTATCTCTGGGCCGGGCACGGGGTGCATTTTCCCGCTTTTTCGCGCACGCTCAAA 412

Qy       2777 GGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGT 2836
Db        413 GGCGGCTACGCCACCATTTCCTCCGAGCGCATGGCCGAGGTGATGAATGAGCGCCTGTGC 472

Qy       2837 GACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCG 2896
Db        473 GCCGCCATCCGCACCAATGCCCGCGTCGCCCATGTGGCGCCCGACCACGTGGTGCTGGAG 532

Qy       2897 GATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCA 2956
Db        533 GGCGGCGAGCGCATCGACGCCCGCGCCGTGGTGGACGGGCGCGGCCCACTCGCCTCGCGC 592

Qy       2957 CATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCAC 3016
Db        593 CATCTGGACCTCGGCTACCAGACCTTCCTGGGGCAGGAACTGCGTATGTCTCGGCCCCAT 652

Qy       3017 GGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTT 3076
Db        653 GGCCTCACCCGCCCCATCATCATGGATGCGCGGGTGGAGCAGCTGGGCGGTTACCGCTTC 712

Qy       3077 GTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAAC 3136
Db        713 GTCTATGTGCTGCCGCTGGACGATGACACGCTGCTGGTGGAAGACACCTATTATGCCGAC 772

Qy       3137 CAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAA 3196
Db        773 GGCCCGGATCTTCCCGCCGATGCCTTGCGCGGCCGCATTTCCGCCTATGCCGCGGCGCAG 832

Qy       3197 GGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGC 3256
Db        833 GGCTGGGCCGTGGATTATGTGGTGCGGGAGGAGGACGGCATCTTGCCGATAGCGTTGGGC 892

Qy       3257 GGCAACATCGATCGATTCTGGCAACAGCAGC---GCGGCCAAGCGTGCAGCGGCCTGCGC 3313
Db        893 GGCGACATCAACGCCTTTCTTGCCGAGACGCCTTCGGGCGTCGCGCCCGCCGGACTGCGG 952

Qy       3314 GCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAG 3373
Db        953 GCAGGCCTGTTCCACCCCACCACCGGCTACAGCCTGCCCGACGCCATGGCCCTGGCCGAC 1012

Qy       3374 TTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTT 3433
Db       1013 TCGGTGAGCGCGCTTGCCGACCTTTCCGGCCCCGCGCTTTCGGCGGCAGTGCGCAGCCAT 1072

Qy       3434 GCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTG 3493
Db       1073 GCGGCGGCGGCGTGGAACGGGCGCGGCTTCTTCCGCCTGCTCAACCGCATGCTGTTCCGC 1132

Qy       3494 GCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGG 3553
Db       1133 GCTGCCGATCCGGAGCGGCGCTATGCCATTCTCCAGCGCTTCTACGGGCTGTCGGAAGAC 1192

Qy       3554 TTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGC 3613
Db       1193 CTGATCGCCCGTTTCTACGCCGATCGCCTCACCCTTGCCGACAAGGCGCGCATCCTTTCG 1252

Qy       3614 GGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCA 3673
Db       1253 GGCCGCCCACCCGTATCGGTGTTCAGGGCGCTCTCCTGCCTTGTTGAGACGAAAGCCGCA 1312

Qy       3674 GGGAAGAAAAAATGAAA------------------------CGCACTTATGTGATTGGCG 3709
Db       1313 CCCGGTTCCCCATGATGCTCGATGCCACCAAGACCATCCGCCGCGCCGCCGTCATCGGGT 1372

Qy       3710 CAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCT 3769
Db       1373 CCGGCTTCGGCGGGCTCTCCCTCGCCATCCGCCTGCAGGCGGCGGGCATCCGCACCACCG 1432

Qy       3770 TACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTA 3829
Db       1433 TGTTCGAGCAGCGGGACAAGCCCGGCGGCCGCGCCTACGTCTATGAGCAGGACGGCTTCA 1492

Qy       3830 CCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGC 3889
Db       1493 CCTTCGATGGCGGCCCCACCGTCATCACCGATCCCACCTGCCTGGAAGAAGTGTTCGCCG 1552

Qy       3890 TGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCC 3949
Db       1553 CCGCCGGCCGCAAGCTGTCCGACTATGTGGAGCTGATGCCGGTCTCGCCGTTCTACCGGC 1612

Qy       3950 TGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGC 4009
Db       1613 TGCTGTGGCCGGACGGCCGGCAGTTCGACTATGTGAACGACCAGGCCGCGCTGGACGCCC 1672

Qy       4010 AGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCAC 4069
Db       1673 AGATCGCCACCTTCGATAAGGCGGACGTGGAGGGCTATCGCCGCTTCCTCGCCTATTCGC 1732

Qy       4070 GTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTG 4129
Db       1733 AGGCGGTCTTCGAGGAGGGTTATCTCAAGCTCGGCGCGGTGCCGTTCCTCGAATTCTCCA 1792

Qy       4130 ACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCA 4189
Db       1793 GCATGATGAAGGCCGCGCCCAAGCTGGTGAGGCTGGAAGCCTGGCGCTCGGTCTATGCCA 1852

Qy       4190 TGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGC 4249
Db       1853 TGGTCTCGCGCTTCATCCGCGACGATCACCTGCGGCAGGCCTTCTCGTTCCATTCCCTGC 1912

Qy       4250 TGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGC 4309
Db       1913 TGGTGGGGGGCAATCCCTTCTCCACCTCGTCCATCTATGCCCTCATCCATGCGCTGGAGC 1972

Qy       4310 GTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGC 4369
Db       1973 GCAAGTGGGGCGTGTTCTTCCCCCGCGGGGAACCGGCGCCTTGGTGCGCGGCATGGTCA 2032

Qy       4370 GACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAA 4429
Db       2033 AGCTGTTCACCGACCTCGGCGGCGAGATCCGCCTCACCAGCCCGGTGGACGAGATCGTGG 2092

Qy       4430 CCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCG 4489
Db       2093 TGGAGGGCCAGCGCGCCACCGCCGTGAAGCTCAAGAGCGGCGAGACCCTGCCGTTCGACC 2152

Qy       4490 CTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGC 4549
Db       2153 TCGTCGCCTCCAATGCGGACGTGGTGCACACCTATCGCCACCTTCTGCGCGGGGCGGCGC 2212

Qy       4550 TGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTAC 4609
Db       2213 GCGGTCGCTCGGAAGGCGCGCGGCTGGCCAAGACGCGCCACTCCATGTCGCTGTTCGTCA 2272

Qy       4610 TCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCC 4669
Db       2273 CCTATTTCGGGGCGCGGCGCACCTGGGACCATCTCCAGCACCACACGGTGCTGTTCGGTC 2332

Qy       4670 CGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTT 4729
Db       2333 CGCGCTATCGGGGCCTCGTGGACGAGATCTTCAAGGGCCCGAACCTTCCCGACGACTTCT 2392

Qy       4730 CACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCA 4789
Db       2393 CGCTCTATCTCCACGCGCCCACCGTGACAGACAAATCGCTGGCGCCGGAAGGCTGTACCG 2452

Qy       4790 GCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGG 4849
Db       2453 CCTTCTACGTGCTCTCGCCGGTGCCGCATCTGGGCAAGGCGGACATCGACTGGGAGGTGG 2512
```

```
Qy     4850 AAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGAT 4909
            |||| |||| ||  ||||| || ||  | ||  |||||||  |||      |||||||
Db     2513 AAGGCCCGCTCTATCGCGACCGCATCCTCGCCCATCTGGAAGAGCGGCTACTGCCGGGGC 2572

Qy     4910 TACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGC 4969
            | ||  |     | || ||     | ||  | || || || |  || ||||||   |||
Db     2573 TGCGCGACAGCCTCGTCACCTCGCGCATCCTCACCCCCTTCGGCTTCCGCGACGAACTGT 2632

Qy     4970 ATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCC 5029
              |||||||| |||||||| || ||| | || ||| |  | || ||    || |||||||
Db     2633 CCGCCCATCAGGGCTCGGCCTTCTCGGTAGAACCGCTCCTCACCCAGTCAGCGTGGTTCC 2692

Qy     5030 GCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATC 5089
            | ||||| |||||||||||| | ||   ||| |||||| | |  || ||||| || ||||
Db     2693 GGCCGCACAACCGCGATGCCAAGATTGCCAACCTCTATTTCGCCGGGGCCGGCACCCATC 2752

Qy     5090 CAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGG 5149
            | || ||||||||||| || ||| | || |||||||||||||| ||  |||| || || |
Db     2753 CGGGGGCGGGCGTGCCGGGGGTGGTGGGCTCGGCCAAGGCCACGGCGGGGCTCATCCTCG 2812

Qy     5150 AGGATCGCGCC---------------GAATGAATCGACAGCCTTTACTTGAGCAAGTAAC 5194
             |||||  | |               | | |  |   |   |  | || |
Db     2813 CGGATCTGGGCGTGACGCCGCGATGAGCACGCCTGCGCCTACCGTCCTCGGTGCCAGCGA 2872

Qy     5195 GCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACC 5254
            | |  ||||      |||||| ||||| |||||| |||||||    ||||| |   | |
Db     2873 GGAGGCCATAAACAAGGGCTCCAAGAGCTTCGCCGCCGCCGCGCGCCTGTTCGGCCCGCG 2932

Qy     5255 GACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGA 5314
             | |||      | || ||||||| ||||| ||||| || |||||||| || ||| | ||
Db     2933 CATGCGGGAAGACGCGGTGATGCTCTATGCCTGGTGCCGCCACTGCGACGACGTGGTGGA 2992

Qy     5315 TGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTAT 5374
             |||||    |||||  | ||  |          |||| |     |||  ||   ||  |
Db     2993 CGGGCAGGAACTGGGGCACGGGCGTGTCTCCAGCGCCGCCTCGCCCGCCGAG---CGGCT 3049

Qy     5375 GCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGC 5434
              |    ||  |    || |||||   ||||||  ||||||||| |   |   | |||||
Db     3050 CGATGGCCTCTATGAGGAGACCCGCAACGCCTATCGCGGCGCGCCCTCCGCCCATCCGGC 3109

Qy     5435 GTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGA 5494
             ||    || ||    || |||| ||      || |  | || | |      |  || |
Db     3110 CTTCGCCGCCTTCGCCGAGGTGGTGAAGCGCAACGACATTCCCGAGCGCTATCCCCTTCA 3169

Qy     5495 TCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACAC 5554
             |  ||||| ||||||   ||||| |       |   |  ||||  | | ||||||||||
Db     3170 ACTGCTGGAGGGCTTCCGCATGGACGTGGAAGGCCGGCGCTACGACACCCTCGATGACAC 3229

Qy     5555 GCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGG 5614
            |||    || ||||| ||||| || |||||||| ||  |||||||||| | | | |||||
Db     3230 GCTCACCTATTGCTACCACGTGGCCGGCGTGGTGGGCGTGATGATGGCCCTCATCATGGG 3289

Qy     5615 CGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCAC 5674
            || ||| || |||| |||||| || |  ||  ||||  | || || |  |||||||||||
Db     3290 CGCGCGGGATGAAGTGGTGCTGGACCGGGCGAGCGACCTCGGCCTCGGCTTCCAGCTCAC 3349

Qy     5675 TAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATC 5734
             ||||| || |||||| |  | || |||||||  |  || |||  ||| ||||||    |
Db     3350 CAACATCGCCCGCGACGTGATGGAGGATGCCGCCATCGGCCGCATCTACCTGCCGGCGGC 3409

Qy     5735 CTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGC 5794
            ||  ||||   |||| ||  | |  |  |   |  |  | || ||| | || || || ||
Db     3410 CTTCCTCGCAGAGGCCGGCGTGCCGGAGGCGGCCATCGCCGCGCCGGAGCACCGCGCGGC 3469

Qy     5795 GCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCG 5854
            | | |||    |||  |||||  |  | |  | ||| || ||||||||  ||  ||| |
Db     3470 GGTGGCCGGCGTGGTGGCGCGGCTGCTCGACGTGGCCGAGCCCTATTACGACCAGGCCCT 3529

Qy     5855 ATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGT 5914
               |||  |  | |   | ||| | ||| |||||   ||  | |   |||| || || ||
Db     3530 CATCGGCATGGCCGCCCTGCCGTTCCGCGCGGCGGCAGCGGTGGGCGCGGCGCGTGGGGT 3589

Qy     5915 TTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCA 5974
             |||||||  || ||| || | || |   |   |||   | | || |||||| | ||
Db     3590 CTATCGCGCCATCGGCATCGAGGTCCGTAAGCGCGGGCCGAAGGCGTGGGATACCCGCGT 3649

Qy     5975 GCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCAC 6034
               ||||   |  |   | ||    |   | || | |   ||    |    |||  || |
Db     3650 CTCCACCTCCACGGCCCAGAAGGCCGGCTTTCTCGCGGGTGGCCTCGCGCAGGCCCTCGC 3709

Qy     6035 TTCGCGTGTGTCTCG---TCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTG 6091
              ||||  |     |    ||    || ||||| ||   ||| |||  ||||||    ||
Db     3710 CACGCGGTTCAGAGGCGCGCCGCCGCCCCGTCCCGCGCATCTTTGGACGCGTCCCGCATG 3769

Qy     6092 A 6092
            |
Db     3770 A 3770


RESULT 11
AF218415
LOCUS       AF218415                6635 bp    DNA     linear   BCT 22-JUN-2000
DEFINITION  Bradyrhizobium sp. ORS278 geranylgeranyl synthase (crtE), lycopene
            cyclase (crtY), phytoene desaturase (crtI), phytoene synthase
            (crtB), and beta-carotene ketolase (crtW) genes, complete cds.
ACCESSION   AF218415
VERSION     AF218415.1  GI:8650414
KEYWORDS    .
SOURCE      Bradyrhizobium sp. ORS278
  ORGANISM  Bradyrhizobium sp. ORS278
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE   1  (bases 1 to 6635)
  AUTHORS   Hannibal,L., Lorquin,J., D'Ortoli,N.A., Garcia,N., Chaintreuil,C.,
            Masson-Boivin,C., Dreyfus,B. and Giraud,E.
  TITLE     Isolation and characterization of canthaxanthin biosynthesis genes
            from the photosynthetic bacterium Bradyrhizobium sp. strain ORS278
  JOURNAL   J. Bacteriol. 182 (13), 3850-3853 (2000)
  MEDLINE   20309720
   PUBMED   10851005
REFERENCE   2  (bases 1 to 6635)
  AUTHORS   Giraud,E. and Angles d'Ortoli,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1999) Laboratoire des Symbioses Tropicales et
            Mediterraneenes, TA/10J, Campus de Baillarguet, Montpellier 34398,
            France
FEATURES             Location/Qualifiers
     source          1. .6635
                     /organism="Bradyrhizobium sp. ORS278"
                     /mol_type="genomic DNA"
                     /strain="ORS278"
                     /db_xref="taxon:114615"
     gene            complement(273. .1223)
                     /gene="crtE"
     CDS             complement(273. .1223)
                     /gene="crtE"
                     /codon_start=1
                     /transl_table=11
                     /product="geranylgeranyl synthase"
                     /protein_id="AAF78199.1"
                     /db_xref="GI:8650415"
                     /translation="MHKPVDLTDTAAFETQLDRWRGRIGEAVAEAMAFGTTVPAPLQA
                     GMSHAVLAGGKRYRGMLVLALGSDLGVPEEQLLSSAVAIETIHAASLVVDDLPCMDDA
                     RRRRSQPATHVAFGEATAILSSIALIARAMEVVARDRQLSPASRSSIVDTLSHAIGPQ
                     ALCGGQYDDLYPPYYATEQDLIHRYQRKTSALFVAAFRCPALLAEVDPETLLRIARAG
                     QRLGVAFQIFDDLLDLTGDAHAIGKDVGQDHGTVTLATLLGPARAAERAADELAAVQK
                     ELRETVGPGRALDLIRRMAARIAGTGKKSAGRDDLRPHAG"
     gene            1446. .2630
                     /gene="crtY"
     CDS             1446. .2630
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /product="lycopene cyclase"
                     /protein_id="AAF78200.1"
                     /db_xref="GI:8650416"
                     /translation="MSRDADVIVIGGGLAGCLIALRLTDARPDLRVVIIEGSASIAGN
                     HTWSFFGTDISSDQHAWLGRLVGHRWPGYEVRFAEHAIRLSTAYLSMTSTRLRAEVEQ
                     RFPERILRDATAISATADHVVLEGGRTLRAPCVIDARGGRPVPGLALGFQKFLGLEVR
                     LAAPHGLDVPIVMDATVAQSDGYRFVYTLPLDPQRLLIEDTYYSDGGELPEQVLHQRI
                     ARYALAKGWQIAEIIRAEQGVLPVILAGDPSGLVSKPDSPPRVGLAALLVHPTTGYSL
                     PDAVRVADLLTARLAQRGALSSADARETIDGYGRTIWRRRGYYRFLNRMLFKAAEPSE
                     RHRILARFYGLDQALIERFYAARIQPQDKLRVFMHMLMKPPIPISSALACLPEASAFK
                     TP"
     gene            2627. .4147
                     /gene="crtI"
     CDS             2627. .4147
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene desaturase"
                     /protein_id="AAF78201.1"
                     /db_xref="GI:8650417"
                     /translation="MSQEMQNAKTAVVIGSGFGGLSLAIRLQSAGIATTLVEKRDKFG
                     GRAYVYEQDGFTFDAGPTVITDPTCLQELFALSGRKLENYVELMPVSPFYQLRWEDGA
                     TFDYVNDQAELERQIAAFCPADVDGYRRFRSYSERLLEEGYVKLGHVPFPDFRSMVRV
                     APQLVALQSYRSVYSKVSQYVSDEHLRQAFSFHSLLVGGNPFATSSIYALIHALERRW
                     GVWFPRGGTGALINKGLVQLFKDLGGEVTLSTSVSRIETANGRVSAVVAEDGRRFAAD
                     IVASNADVVHTYRDLLKDEPLARPTAQSLMRKRFSMSLFVIYFGLRREHPELKHHIIL
                     FGRRYRELINEIFKGPALPEDFSLYLHAPSVTDPSLAPQGCSTYYVLSPVPHLAAAPI
                     DWSVEGPRYRDRILDYLEARILPGLKSDLATCRIFTPQDFNTELNAHLGSAFSLEPIL
                     TQSAYFRAHNADDKIKGLYLVGAGTHPGAGIPGVVGSAKATARVILEDQRELAEAQ"
     gene            4144. .5151
                     /gene="crtB"
     CDS             4144. .5151
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene synthase"
                     /protein_id="AAF78202.1"
                     /db_xref="GI:8650418"
                     /translation="MTLRDNDHLARLSEAVIRNGSKSFAAASKLFDSRTRASVHLLYA
                     WCRHCDDVIDGQDLGIRQGVGAPGPQIGTLQMLRDQTAQALEGAPMRDPVFQGLQRVV
                     QEHAIPHHHVFELLDGFAMDVDGREYETLSETLDYCYHVAGVVGVMMSAIMGAREEAT
                     LDRAADLGIALQLTNIARDVIEDAQTGRMYLPQQWLCEAGVPAAEVAEPQHRQAVFRV
                     VARLLDVAEQFYEASEAGIARLPVRCAWAVETARVVYRQIGREVMKRGPGAWDARIAT
                     TGAQKLGAIGRSALTLGLTRTWVKVPAREHNLWTRPKARRAASDAQASFANACSNRQG
                     Y"
     gene            5123. .5899
                     /gene="crtW"
     CDS             5123. .5899
                     /gene="crtW"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-carotene ketolase"
                     /protein_id="AAF78203.1"
                     /db_xref="GI:8650419"
                     /translation="MHAATAKATEFGASRRDDARQRRVGLTLAAVIIAAWLVLHVGLM
                     FFWPLTLHSLLPALPLVVLQTWLYVGLFIIAHDCMHGSLVPFKPQVNRRIGQLCLFLY
                     AGFSFDALNVEHHKHHRHPGTAEDPDFDEVPPHGFWHWFASFFLHYFGWKQVAIIAAV
                     SLVYQLVFAVPLQNILLFWALPGLLSALQLFTFGTYLPHKPATQPFADRHNARTSEFP
                     AWLSLLTCFHFGFHHEHHLHPDAPWWRLPEIKRRALERRD"
ORIGIN

  Query Match             15.8%;  Score 1105.8;  DB 1;  Length 6635;
  Best Local Similarity   58.4%;  Pred. No. 5.6e-222;
  Matches 2082;  Conservative    0;  Mismatches 1417;  Indels   66;  Gaps    6;

Qy        2502 GTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGAC 2561
               | ||  || | |  | | | ||| |     |   ||| || ||  |  |||| |  || |
Db        1426 GAGCGCGCTGCGTTGTCGTCATGAGTCGAGATGCCGACGTCATCGTCATCGGCGGCGGTC 1485

Qy        2562 TGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGT 2621
               | ||     || |||| ||  |||| ||     |    | |||  | |||     | |
Db        1486 TTGCCGGCTGCCTGATCGCCTTGCGGCTCACCGACGCACGGCCGGATCTGCGCGTCGTCA 1545

Qy        2622 TGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATC 2681
               |  | ||  ||   ||       || |||||||| ||||||   ||   |   | ||||
Db        1546 TCATCGAAGGCTCGGCCAGCATCGCCGGCAATCACACCTGGAGCTTCTTTGGAACCGATA 1605

Qy        2682 TCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATC 2741
               ||    |||| || |    ||||||   || |||  |       || ||| | ||||||
Db        1606 TCTCGTCCGACCAGCACGCCTGGCTCGGACGGCTCGTCGGTCATCGCTGGCCGGGTTATG 1665

Qy        2742 AGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAG 2801
               ||||||| ||  |  |  |    |  |  ||||      |  ||    |||||  | ||
Db        1666 AGGTGCGCTTCGCCGAACATGCCATCCGTCTGTCGACCGCCTACCTCTCCATGACGTCGA 1725

Qy        2802 GCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCG 2861
                     |      | |  |  |   |    |    ||     |   |    || |  | |
Db        1726 CGCGGCTGCGTGCCGAGGTCGAGCAGCGTTTTCCGGAGAGGATCCTGCGCGACGCGACGG 1785

Qy        2862 TACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAG 2921
                       | ||   |    || ||   ||||| ||  |||||     |    ||||
Db        1786 CCATCTCCGCAACGGCAGACCATGTCGTGCTGGAGGGCGGCCGCACCTTGCGCGCGCCCT 1845

Qy        2922 TGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGT 2981
                 ||||| ||||  ||||||   | ||||  |||    ||    ||  || | ||||  ||
Db        1846 GCGTGATCGATGCGCGCGGCGGCCGGCCGGTGCCGGGGCTCGCTCTCGGGTTTCAGAAGT 1905

Qy        2982 TTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGG 3041
               | || || |  |||  || ||||||   ||||||||||||  |   ||| ||| ||||||
Db        1906 TCCTCGGGCTCGAGGTGCGGCTGGCCGCGCCGCACGGCCTCGATGTGCCCATCGTGATGG 1965

Qy        3042 ATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCG 3101
               | || || ||||  |||   |  || ||||| || ||||||||||||||||||  | |
Db        1966 ACGCGACGGTCGCGCAGAGCGACGGCTATCGCTTCGTCTACACGCTGCCGCTCGACCCTC 2025

Qy        3102 ATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCG 3161
               | ||||| ||||| || || ||| | ||     ||     || | || | ||| |
Db        2026 AGCGGCTGTTGATCGAGGACACCTACTATAGCGACGGCGGCGAGTTGCCCGAGCAGGTGC 2085

Qy        3162 CTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGC 3221
                 | ||||| ||||||   ||| ||        | ||||||  | |      | |  | |
Db        2086 TGCATCAGCGCATCGCGCGCTACGCGCTCGCCAAGGGCTGGCAGATCGCCGAGATCATCC 2145

Qy        3222 GTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGCAAC 3281
               | |  ||||| ||  |  | ||  | |  |||   ||| |        |  |
Db        2146 GCGCGGAGCAGGGTGTGCTGCCCGTCATTCTGGCGGGCGATCCGTCCGGGCTGGTCAGCA 2205

Qy        3282 AGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTT 3341
               |||    | |||   || ||   || ||    ||    ||  | ||| | || ||||| |
Db        2206 AGCCCGACAGCCCGCCGCGCGTGGGTCTCGCGGCGCTTCTCGTGCATCCGACGACCGGCT 2265

Qy        3342 ACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTG-TTGCCCACCGAT 3400
               | ||  |||||  ||| ||     | || ||  || |  | |||| | |||| || ||
Db        2266 ATTCGCTGCCGGACGCGGTCCGCGTCGCCGATCTGCTGACGGCGCGGCTTGCGCAACGGG 2325

Qy        3401 GCCCTCACGCTCA-----------GCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGC 3449
               || | | | |              || |    |||||  |||  | |||   |   ||||
Db        2326 GCGCGCTCTCGAGCGCTGACGCGCGCGAGACGATCGATGGCTACGGCCGGACGATCTGGC 2385

Qy        3450 GCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGC 3509
               ||   |   | |  |  ||  | || ||||| ||||||||   |||||   ||||
Db        2386 GCCGGCGCGGCTACTATCGCTTCCTGAACCGGATGCTGTTCAAGGCCGCCGAGCCCTCCG 2445

Qy        3510 AGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTT 3569
               ||||    ||  |  ||  ||| |||||| | ||||||   |   | ||    ||||| |
Db        2446 AGCGTCACCGAATCCTGGCACGATTTTACGGTCTCGATCAGGCCCTGATCGAGCGCTTCT 2505

Qy        3570 ACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT---------TCTGTGCGGCAAGC 3620
               ||||||  |   | |  | ||  || ||   |||  |         | ||      ||||
Db        2506 ACGCCGCCCGGATCCAGCCGCAGGACAAGCTGCGCGTCTTCATGCATATGCTGATGAAGC 2565

Qy        3621 CGCCGGTGCCCATCGGTGAAGCGCT------------------GCGCGCGCTGTTGAATT 3662
               ||||| | || |||      |||||                  | ||||  |   ||
Db        2566 CGCCGATCCCGATCTCGTCCGCGCTCGCCTGCCTGCCTGAGGCGAGCGCCTTCAAGACCC 2625

Qy        3663 CTGTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGG 3722
               |    |      |    | ||    ||     |    ||||| ||  | || ||||| ||
Db        2626 CATGAGTCAAGAGATGCAGAACGCCAAGACAGCAGTGGTGATCGGTTCGGGATTTGGTGG 2685
```

```
Qy       3723 CCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCG 3782
               ||  |||| || || |||||||||| |||| || ||  | || ||  |  ||||| ||||
Db       2686 GCTTTCGCTCGCCATACGCCTGCAATCGGCCGGGATCGCGACAACGCTCGTCGAGAAGCG 2745

Qy       3783 CGACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGG 3842
              ||||||    ||||| ||||||||||| |  ||| |    ||||| |||||||| |||||
Db       2746 CGACAAGTTCGGCGGCCGCGCCTATGTCTACGAGCAGGACGGCTTCACCTTCGACGCCGG 2805

Qy       3843 ACCCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATC 3902
               ||||| || ||||||||||| | |  | |  | ||| ||||| ||||  | ||
Db       2806 GCCCACTGTCATCACCGATCCGACCTGCCTGCAGGAGCTGTTCGCGCTCTCGGGGCGCAA 2865

Qy       3903 GCTCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGA 3962
              ||||    | || |||||||||||||||||  | || |||||||  ||| ||||||| ||
Db       2866 GCTCGAGAACTATGTCGAGCTGATGCCGGTCTCCCCATTCTATCAGCTGCGCTGGGAGGA 2925

Qy       3963 TGGCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTT 4022
               |||       | || |||| |||  | ||| |   ||||||||  ||||||||  ||||
Db       2926 CGGCGCCACTTTCGACTACGTCAACGACCAGGCCGAGCTGGAGCGCCAGATCGCGGCGTT 2985

Qy       4023 CAATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAG 4082
              |  ||||   ||||| || |||||||| |  || |   | |||          |  |
Db       2986 CTGTCCGGCTGATGTGGACGGCTATCGGCGCTTCCGGTCGTATAGCGAGCGCCTGCTGGA 3045

Qy       4083 AGAGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGT 4142
               ||||| ||  | || ||||||   |||||||| | | |  | ||   |||| |||| ||
Db       3046 GGAGGGCTACGTCAAGCTCGGCCACGTGCCGTTCCCGGATTTCCGGAGCATGGTGCGTGT 3105

Qy       4143 CGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATT 4202
               ||||||||| |||     |||||    |  ||||||||||| |||| |||| || |||
Db       3106 GGCGCCGCAGCTGGTGGCGCTGCAGAGCTATCGCAGCGTCTATAGCAAGGTGTCGCAATA 3165

Qy       4203 TATTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAA 4262
                |    ||||| |||||||| ||||| ||   |||||||||  ||||||| ||||| ||
Db       3166 CGTCTCCGACGAGCATCTGCGGCAGGCCTTCAGCTTCCACTCGCTGCTGGTCGGCGGCAA 3225

Qy       4263 TCCTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGT 4322
              ||| || || ||||| ||||||||  |  | || ||||||||||||||    ||||| ||
Db       3226 TCCGTTCGCGACGTCCTCGATCTACGCGCTGATCCATGCGCTGGAGCGCCGCTGGGGAGT 3285

Qy       4323 GTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGT---GCAGGGCATGGCGCGACTGTTCGA 4379
               ||||| |||||||||||||||||||||||  |     ||||  | |  |  |||||| |
Db       3286 CTGGTTCCCGCGCGGCGGCACCGGCGCGCTCATCAACAAGGGGCTCGTCCAGCTGTTCAA 3345

Qy       4380 GGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAA 4439
              |||  | ||||||||| ||   |||    ||   ||||||| | | || ||      |
Db       3346 GGATCTCGGCGGCGAGGTGACGCTGTCGACCAGCGTGAGCCGGATCGAGACGGCGAACGG 3405

Qy       4440 TCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTC 4499
               ||  | || | |||       |||| ||| || || ||||  |||| |   || |||
Db       3406 GCGGGTCAGTGCCGTGGTCGCCGAGGACGGCCGTCGGTTCGCCGCCGACATCGTCGCCAG 3465

Qy       4500 CAATGCCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAA 4559
              ||| |||||||| || |||||||||  | | || ||     || | ||||||||   |
Db       3466 CAACGCCGACGTCGTTCATACCTACCGCGATCTCCTGAAGGACGAGCCGCTGGCGCGGCC 3525

Qy       4560 ACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGG 4619
                  |||  |||  ||| ||| |||||| | ||||  |||||||| ||  ||||||| ||
Db       3526 GACGGCGCAATCTTTGATGCGCAAGCGCTTCAGCATGTCGCTGTTCGTGATCTATTTCGG 3585

Qy       4620 CCTGAATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCG 4679
              ||||   |    ||||    |||||  |||||||| | |    || |||| ||| || ||
Db       3586 CCTGCGCCGCGAGCATCCGGAGCTCAAGCACCACATCATTCTGTTCGGCCGGCGCTACCG 3645

Qy       4680 TGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCT 4739
               ||| ||||| | ||||| |||||  |   |  | || | || ||||| || || || ||
Db       3646 CGAGCTGATCAACGAGATCTTCAAGGGGCCCGCGTTGCCTGAGGATTTCTCGCTGTATCT 3705

Qy       4740 GCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGT 4799
              |||||||||  ||   | |||||||||||| || || |  |||||| ||| || ||| ||
Db       3706 GCACGCGCCGAGCGTGACCGATCCGTCGCTCGCGCCACAAGGCTGCAGCACCTATTACGT 3765

Qy       4800 GTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCG 4859
              | |  |||||||||||||||||| | ||||    |||||||||     ||| || |||||
Db       3766 GCTCTCGCCGGTGCCGCATCTCGCCGCCGCACCGATCGACTGGAGCGTGGAGGGCCCGCG 3825

Qy       4860 CTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCA 4919
              ||  ||||| || ||  | |  ||||| |||  || |    |||||||  |      | |
Db       3826 CTATCGCGACCGCATCCTCGACTATCTCGAGGCGCGCATTCTGCCGGGGCTGAAGTCGGA 3885

Qy       4920 ATTAGTGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCA 4979
                | |  ||   | | || || ||||||   |||||   |  |  |||| |||| ||||
Db       3886 CCTCGCCACCTGCCGCATCTTCACGCCGCAGGATTTCAACACCGAGCTGAATGCGCATCT 3945

Qy       4980 CGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAA 5039
              |||||| ||||| ||||||||||||||  ||||||| ||||| |  |||||| | || ||
Db       3946 CGGCTCCGCGTTCTCGCTGGAGCCGATCCTGACGCAGAGCGCGTATTTCCGCGCCCACAA 4005

Qy       5040 CCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGG 5099
              |   || | | | ||||     || ||||| || || ||||| || ||||| ||||| ||
Db       4006 CGCGGACGACAAGATCAAGGGGCTGTATCTCGTCGGCGCCGGCACCCATCCGGGCGCCGG 4065

Qy       5100 CGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGC 5159
              | | ||||||||  |||| ||||||||||| |||||  |  | || || ||||| |
Db       4066 CATTCCCGGCGTCGTCGGCTCGGCCAAGGCGACCGCGCGAGTCATCCTCGAGGACCAGCG 4125

Qy       5160 CGAATGAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTG--------- 5210
              ||||      ||    |  |      ||  |  |   |     |||||  |
Db       4126 CGAACTCGCGGAGGCGCAATGACGCTGCGCGACAATGACCATCTGGCGCGGCTGAGCGAA 4185

Qy       5211 ---------------GGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCG 5255
                             |||||||||||||||||  ||||  |||||||||| ||| | |
Db       4186 GCCGTCATCCGCAACGGCTCGAAGAGTTTCGCAGCCGCATCCAAGCTGTTCGATTCCCGC 4245

Qy       5256 ACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGAT 5315
              || ||   ||||  ||   |||| || || ||||| || || ||||| || || || ||
Db       4246 ACCCGTGCCAGCGTGCACCTGCTCTACGCCTGGTGCCGGCATTGCGACGACGTCATCGAC 4305

Qy       5316 GGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATG 5375
              || ||    || ||     |             ||    |       |   |    | ||
Db       4306 GGACAGGATCTCGGAATTCGGCAGGGCGTCGGCGCGCCTGGCCCGCAGATCGGGACTTTG 4365

Qy       5376 CAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCG 5435
              |||   ||||      | ||    |  ||       || ||||  |||   || |||| |
Db       4366 CAGATGCTGCGCGACCAGACGGCGCAGGCGCTGGAGGGGGCGCCGATGCGCGATCCGGTG 4425

Qy       5436 TTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGAT 5495
              ||   ||  || |||   || | |      |||  | | ||||| || |  | ||| ||
Db       4426 TTCCAGGGATTGCAGCGCGTCGTGCAGGAGCACGCGATTCCGCACCATCACGTGTTCGAG 4485

Qy       5496 CATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACG 5555
              |  || || |||||||| |||||||    |       | |||| ||   |    || |||
Db       4486 CTGCTCGACGGCTTCGCCATGGATGTCGACGGGCGCGAATACGAGACGCTGAGCGAGACG 4545

Qy       5556 CTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGC 5615
              |||   |||||||||||||| || |||||||||||  |||||||| ||    | ||||||
Db       4546 CTGGACTACTGCTATCACGTGGCCGGCGTGGTCGGCGTGATGATGTCGGCGATCATGGGC 4605

Qy       5616 GTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACT 5675
              |  ||||| || |||  ||| || | |||   |||| | ||  | ||| | ||||||||
Db       4606 GCTCGCGAGGAGGCGACGCTGGACCGCGCGGCCGATCTCGGCATCGCGCTGCAGCTCACC 4665

Qy       5676 AACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCC 5735
              ||||| || || ||| |  | || |||||| | |  || |||   |||||||||||
Db       4666 AACATCGCCCGTGACGTGATCGAGGATGCCCAGACCGGCCGCATGTATCTGCCGCAGCAA 4725

Qy       5736 TGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCG 5795
              |||||     |||| ||  | |  ||||    | |  | | |||||| |||||    |||
Db       4726 TGGCTGTGCGAGGCCGGCGTGCCGGCCGCCGAGGTCGCGGAACCGCAGCATCGGCAGGCG 4785

Qy       5796 CTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGA 5855
               |   |    | |  |||||  |  | |  |  ||||| |  | |||  |  |
Db       4786 GTGTTCCGTGTCGTCGCGCGGCTGCTCGATGTCGCGGAGCAGTTTTACGAGGCCAGCGAG 4845

Qy       5856 TCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTT 5915
               | ||  |  |  |  | ||| |||| |  || |||||  | |  ||||| |||| |||
Db        4846 GCAGGCATCGCCCGGCTGCCGGTGCGTTGCGCCTGGGCGGTGGAGACGGCCCGCGTCGTC 4905

Qy        5916 TATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAG 5975
               |||||| | || ||   | | ||   | |   ||||     |||||||||  | ||
Db        4906 TATCGCCAGATCGGGCGCGAGGTGATGAAGCGCGGTCCCGGCGCCTGGGACGCCCGCATC 4965

Qy        5976 CGCACCAGTAAAGGTGAAAAACTGG 6000
                 ||| |     |   | || || |
Db        4966 GCCACGACAGGCGCGCAGAAGCTCG 4990


RESULT 12
PMCRTWZYI
LOCUS       PMCRTWZYI               6335 bp    DNA     linear   BCT 15-SEP-1999
DEFINITION  Paracoccus marcusii crtW, crtZ, crtY, crtI, crtB & crtE genes.
ACCESSION   Y15112
VERSION     Y15112.1  GI:5912291
KEYWORDS    astaxanthin biosynthesis; beta-carotene C-4 oxygenase; carotene
            hydroxylase; crtB gene; crtE gene; crtI gene; crtW gene; crtY gene;
            crtZ gene; geranylgeranyl pyrophosphate synthase; ketolase;
            lycopene beta-cyclase; phytoene desaturase; phytoene synthase.
SOURCE      Paracoccus marcusii
  ORGANISM  Paracoccus marcusii
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
            Rhodobacteraceae; Paracoccus.
REFERENCE   1
  AUTHORS   Harker,M. and Hirschberg,J.
  TITLE     Carotenoid biosynthesis genes in the bacterium Paracoccus marcusii
            MH1
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6335)
  AUTHORS   Hirschberg,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-1997) J. Hirschberg, The Hebrew Univ of
            Jerusalem, Dept of Genetics, Life Sciences Institute, Jerusalem
            91904, ISRAEL
FEATURES             Location/Qualifiers
     source          1. .6335
                     /organism="Paracoccus marcusii"
                     /mol_type="genomic DNA"
                     /strain="MH1"
                     /db_xref="taxon:59779"
     gene            409. .1137
                     /gene="crtW"
     CDS             409. .1137
                     /gene="crtW"
                     /function="astaxanthin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="beta-carotene C-4-oxygenase (ketolase)"
                     /protein_id="CAB56059.1"
                     /db_xref="GI:5912292"
                     /db_xref="GOA:Q9RLH7"
                     /db_xref="TrEMBL:Q9RLH7"
                     /translation="MSAHALPKADLTATSLIVSGGIIAAWLALHVHALWFLDAAAHPI
                     LAVANFLGLTWLSVGLFIIAHDAMHGSVVPGRPRANAAMGQLVLWLYAGFSWRKMIVK
                     HMAHHRHAGTDDDPDFDHGGPVRWYARFIGTYFGWREGLLLPVIVTVYALILGDRWMY
                     VVFWPLPSILASIQLFVFGTWLPHRPGHDAFPDRHNARSSRISDPVSLLTCFHFGGYH
                     HEHHLHPTVPWWRLPSTRTKGDTA"
     gene            1134. .1622
                     /gene="crtZ"
     CDS             1134. .1622
                     /gene="crtZ"
                     /function="astaxanthin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="carotene hydroxylase"
                     /protein_id="CAB56060.1"
                     /db_xref="GI:5912293"
                     /db_xref="GOA:Q9RLH6"
                     /db_xref="TrEMBL:Q9RLH6"
                     /translation="MTNFLIVVATVLVMELTAYSVHRWIMHGPLGWGWHKSHHEEHDH
                     ALEKNDLYGLVFAVIATVLFTVGWIWAPVLWWIALGMTVYGLIYFVLHDGVVHQRWPF
```

```
                     RYIPRKGYARRLYQAHRLHHAVEGRDHCVSFGFIYAPPVDKLKQDLKTSGVLRAEAQE
                     RT"
     gene            1619. .2779
                     /gene="crtY"
     CDS             1619. .2779
                     /gene="crtY"
                     /function="astaxanthin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="lycopene beta-cyclase"
                     /protein_id="CAB56061.1"
                     /db_xref="GI:5912294"
                     /db_xref="TrEMBL:Q9RLH5"
                     /translation="MTHDVLLAGAGLANGLIALALRAARPDLRVLLLDHAAGPSEGHT
                     WSCHDPDLSPHWLARLKPLRRANWPDQEVRFPRHARRLATGYGSLDGAALADAVVRSG
                     AEIRWDSDIALLDAQGATLSCGTRIEAGSVLDGRGPHPSRHLTLGFHKFLGVEIETDR
                     PHGVPRPVIMDGTVTQRDGYGFIYLLPFSRTRILIEDTRYSDGGDLDDDALAAASQDY
                     ARQQGWTGAEVRRERGILPIALAHDAAGFWADHAEGPVPVGLRAGFFHPVTGYSLPYA
                     AQVADVVAGLSGPPGTDALRGAIRDYAIDRARRDRFLRLLNRMLFRGCAPDRRYTLLQ
                     RLYRMPHGLIERFYAGRLSVADQLRIVTGKPPIPLGTAIRCLPERPLLKENA"
     gene            2776. .4281
                     /gene="crtI"
     CDS             2776. .4281
                     /gene="crtI"
                     /function="astaxanthin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene desaturase"
                     /protein_id="CAB56062.1"
                     /db_xref="GI:5912295"
                     /db_xref="GOA:Q9RLH4"
                     /db_xref="TrEMBL:Q9RLH4"
                     /translation="MNAHSPAAKTAIVIGAGFGGLALAIRLQSAGIATTLVEARDKPG
                     GRAYVWHDQGHVFDAGPTVITDPDALKELWALTGQDMARDVTLMPVSPFYRLMWPGGK
                     VFDYVNEADQLERQIAQFNPDDLEGYRRFRDYAEEVYQEGYVKLGTVPFLKLGQMLKA
                     APALMKLEAYNSVHAKVATFIKDPYLRQAFSYHTLLVGGNPFSTSSIYALIHALERRG
                     GVWFAKGGTNQLVAGMVALFERPGGQMMLNAKVARIETEGARTTGVTLADGRSLRADM
                     VASNGDVMHNYRDLLGHTARGQSRAKSLDRKRWSMSLFVLHFGLREAPKDIAHHTILF
                     GPRYKELVNEIFKGPKLAEDFSLYXHSPCTTDPDMAPPGMSTHYVLAPVPHLSRAEID
                     WAVEGPRYADRILAFLEERLIPNLRANLTRTRIFTPADFASELNAHHGSAFSVEPILT
                     QSAWFRPHNRDKTIRNFYLVGAGTHPGAGIPGVVGSAKATAQVMLSDLASA"
     gene            4278. .5192
                     /gene="crtB"
     CDS             4278. .5192
                     /gene="crtB"
                     /function="astaxanthin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene synthase"
                     /protein_id="CAB56063.1"
                     /db_xref="GI:5912296"
                     /db_xref="GOA:Q9RLH3"
                     /db_xref="TrEMBL:Q9RLH3"
                     /translation="MSDLVLTSTEAITQGSQSFATAAKLMPPGIRDDTVMLYAWCRHA
                     DDVIDGQALGSRPEAVNDPQARLDGLRADTLAALQGDGPVTPPFAALRAVARRHDFPQ
                     AWPMDLIEGFAIDVEARDYRTLDDVLEYSYHVAGILGVMMARVMGVRDHPVLDRACDL
                     GLAFQLTHIARDVIDYARIGRCYLPGDWLDHAGARVDGPVPSPELYTVILRLLDAAEP
                     YYPSARVGLADLPPRCAWSIPAAFRIYRAIGLPIRKGGPEAYRQRISTSKAAKIGLLG
                     IGGWDVADHACRGSGVSRQDLWTRPHHA"
     gene            complement(5189. .6064)
                     /gene="crtE"
     CDS             complement(5189. .6064)
                     /gene="crtE"
                     /function="astaxanthin biosynthesis"
                     /codon_start=1
                     /transl_table=11
                     /product="geranylgeranyl pyrophosphate synthase"
                     /protein_id="CAB56064.1"
                     /db_xref="GI:5912297"
                     /db_xref="GOA:Q9RLH2"
                     /db_xref="TrEMBL:Q9RLH2"
                     /translation="MRRDVNPIHATLLQTRIEEIAQGFGAVLQPLGAAMAALSSGKRF
                     RGMLMLLAAEASGGVCDTIVDAACAVEMVHAASLIFDDLPCMDDAGLRRGQPATHVAH
```

```
                    GESRAVLGGIALITEAMALLAGARGASGTVRAQLVRILSRSLGPQGLCAGQDLDLHAA
                    KNGAGVEQEQDLKTGVLFIAGLEMLAVIKEFDAEEQTQMIDFGRQLGRVFQSYDDLLE
                    VVGDQAALGKDTGRDAAAPGPRRGLLAVSDLQKVSRQYEASRAQLEGMVGSKRLQAPE
                    IEALLERVLPYAARA"
ORIGIN

  Query Match             14.9%;  Score 1041.2;  DB 1;  Length 6335;
  Best Local Similarity   57.4%;  Pred. No. 2.3e-208;
  Matches 2059;  Conservative    1;  Mismatches 1479;  Indels   51;  Gaps    8;

Qy       2456 GCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGC 2515

Db       1547 GCCCCCGGTCGACAAGCTGAAGCAGGACCTGAAGACGTCGGGCGTACTGCGGGCCGAGGC 1606

Qy       2516 GACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTG 2575

Db       1607 GCAGGAGCGCACGTGACCCATGACGTGCTGCTGGCAGGGGCGGGCCTTGCGAACGGGCTG 1666

Qy       2576 ATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGAT 2635

Db       1667 ATCGCCCTGGCGCTGCGCGCGGCGCGGCCCGACCTGCGGGTGCTGCTGCTGGACCATGCG 1726

Qy       2636 GCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAA 2695

Db       1727 GCGGGACCGTCAGAAGGCCATACCTGGTCATGCCACGACCCCGATCTGTCGCCGCACTGG 1786

Qy       2696 CTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCT 2755

Db       1787 CTGGCGCGGCTGAAGCCCCTGCGCCGCGCCAACTGGCCCGACCAGGAGGTGCGCTTTCCC 1846

Qy       2756 GCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGC 2815

Db       1847 CGCCATGCCCGGCGGCTGGCCACCGGTTACGGGTCGCTGGACGGGGCGGCGCTGGC---G 1903

Qy       2816 CATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAA 2875

Db       1904 GATGCGGTGGTCCGGTCGGGCGCCGAGATCCGCTGGGACAGCGACATCGCCTTGCTGGAT 1963

Qy       2876 CCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGT 2935

Db       1964 GCGCAGGGGGCGACACTGTCCTGCGGCACCCGGATCGAGGCCGGCTCGGTCCTGGACGGG 2023

Qy       2936 CGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAG 2995

Db       2024 CGGGGCCCGCACCCGTCCCGGCATCTGACCTTGGGTTTCCACAAATTCCTGGGTGTCGAG 2083

Qy       2996 TGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGAT 3055

Db       2084 ATCGAGACCGACCGCCCCCACGGCGTGCCCCGCCCGGTGATCATGGACGGGACAGTCACC 2143

Qy       3056 CAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATT 3115

Db       2144 CAGCGGGACGGGTACGGCTTCATTTATCTGCTGCCCTTCTCTCGGACGCGCATCCTGATC 2203

Qy       3116 GAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATC 3175

Db       2204 GAGGACACGCGCTATTCCGATGGCGGCGATCTGGACGACGACGCGCTGGCGGCGGCGTCC 2263

Qy       3176 GCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGC 3235

Db       2264 CAAGACTATGCCCGCCAGCAGGGTTGGACCGGGGCCGAGGTCCGGCGCGAA---CGGGGC 2320

Qy       3236 ATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGG---CAACAGCAGCGCGGC 3292

Db       2321 ATCCTGCCTATTGCGTTGGCCCATGACGCGGCGGGCTTCTGGGCCGATCACGCGGAGGGG 2380

Qy       3293 CAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCG 3352

Db       2381 CCTGTTCCCGTGGGACTGCGCGCGGGGTTCTTTCACCCGGTCACCGGCTATTCGCTGCCC 2440

Qy       3353 TCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGT---TGCCCACCGATGCCCTCACG 3409

Db       2441 TATGCGGCGCAGGTGGCGGACGTGGTGGCGGGCCTGTCCGGGCCGCCCGGCACCGACGCG 2500

Qy       3410 CTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGT 3469

Db       2501 CTGCGCGGCGCCATCCGCGATTACGCGATCGACCGGGCACGCCGCGACCGCTTTCTGCGC 2560

Qy       3470 CTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAA 3529

Db       2561 CTGCTGAACCGGATGCTGTTCCGCGGCTGCGCGCCCGACCGGCGCTATACCCTGCTGCAG 2620

Qy       3530 CGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTG 3589

Db       2621 CGGTTATACCGCATGCCGCATGGACTGATCGAACGGTTCTATGCCGGCCGGCTGAGCGTG 2680

Qy       3590 CGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGC 3649

Db       2681 GCGGATCAGCTGCGCATCGTGACCGGCAAGCCTCCCATTCCCCTTGGCACGGCCATACGC 2740

Qy       3650 GCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATG--------------------A 3688

Db       2741 TGCCTGCCCGAACGTCCCCTGCTGAAGGAAAACGCATGAACGCCCATTCGCCCGCGGCCA 2800

Qy       3689 AACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAG 3748

Db       2801 AAACCGCCATCGTGATCGGCGCAGGCTTTGGCGGGCTGGCCCTGGCCATCCGCCTGCAGT 2860

Qy       3749 CGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATG 3808

Db       2861 CCGCGGGCATCGCCACCACCCTGGTCGAGGCCCGCGACAAGCCCGGCGGGCGCGCCTATG 2920

Qy       3809 TGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCG 3868

Db       2921 TCTGGCACGATCAGGGCCATGTCTTCGACGCGGGCCCGACTGTCATCACCGACCCCGATG 2980

Qy       3869 CCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGC 3928

Db       2981 CGCTCAAGGAGCTGTGGGCGCTGACCGGGCAGGACATGGCGCGCGACGTGACGCTGATGC 3040

Qy       3929 CGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATA 3988

Db       3041 CGGTGTCGCCCTTCTATCGGCTGATGTGGCCGGGCGGAAGGTCTTCGATTACGTGAACG 3100

Qy       3989 ATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATC 4048

Db       3101 AGGCCGATCAGCTGGAACGCCAAATCGCCCAGTTCAACCCGGACGACCTGGAAGGATACC 3160

Qy       4049 GTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGG 4108

Db       3161 GTCGCTTCCGTGATTACGCTGAGGAGGTGTACCAGGAGGGCTACGTCAAGCTGGGCACCG 3220

Qy       4109 TGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAG 4168

Db       3221 TGCCCTTCCTCAAGCTGGGCCAAATGCTCAAGGCCGCGCCCGCGCTGATGAAGCTTGAGG 3280

Qy       4169 CATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGG 4228

Db       3281 CCTACAATTCCGTCCATGCCAAGGTCGCGACCTTCATCAAGGACCCCTATCTGCGGCAGG 3340

Qy       4229 CGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATA 4288

Db       3341 CCTTTTCGTATCACACGCTGCTGGTGGGCGGAATCCCTTCTCGACCAGCTCGATCTATG 3400

Qy       4289 CCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCG 4348

Db       3401 CGCTGATCCACGCGCTGGAGCGGCGCGGCGGGGTCTGGTTCGCCAAGGGCGGCACCAACC 3460

Qy       4349 CGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATG 4408

Db       3461 AGCTGGTCGCGGGCATGGTCGCGCTGTTCGAACGGCCTGGCGGCCAGATGATGCTGAACG 3520

Qy       4409 CCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCG 4468

Db       3521 CCAAGGTCGCCCGGATCGAGACCGAGGGCGCGCGGACCACGGGCGTCACCCTGGCCGACG 3580
```

```
Qy        4469 GACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACA 4528
               | ||    ||    |||| |   || |||  ||| | ||||||  |||| | |||   |
Db        3581 GGCGGTCTTTGAGGGCCGACATGGTCGCCAGCAACGGCGACGTCATGCACAACTATCGCG 3640

Qy        4529 AACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCA 4588
               | ||| |  |||||    | |  |    ||  |     | ||||||| | || ||||||
Db        3641 ACCTGTTGGGCCACACGGCCCGCGGGCAGAGCCGCGCGAAATCGCTGGACCGCAAGCGCT 3700

Qy        4589 TGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGC 4648
                |  ||  | | |||| || ||  |||| || |||    || |||   |  |  ||||||
Db        3701 GGTCCATGTCGTTGTTCGTGCTGCATTTCGGTCTGCGCGAGGCGCCCAAGGACATCGCGC 3760

Qy        4649 ACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCA 4708
               | |||||| ||   || |||||| || ||    ||| || || | ||||| |||||  ||
Db        3761 ATCACACCATCCTGTTCGGCCCCCGCTACAAGGAGCTGGTCAACGAGATATTCAAGGGCC 3820

Qy        4709 GCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGC 4768
                  ||||||| || ||||| || || || |:|||  ||||||||| || |||||||
Db        3821 CGAAGCTGGCCGAGGATTTCTCGCTGTATCYGCATTCGCCCTGCACCACCGATCCGGACA 3880

Qy        4769 TGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCG 4828
               |||| || || |||     ||    ||| ||| | || ||||||||||||||  ||  ||
Db        3881 TGGCGCCTCCGGGCATGTCCACGCATTACGTGCTGGCCCCGGTGCCGCATCTGAGCCGCG 3940

Qy        4829 CTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGG 4888
               | || ||||| |||      || || |||||||    ||| || ||  | || |  ||||
Db        3941 CCGAGATCGATTGGGCGGTCGAGGGGCCGCGCTATGCCGACCGCATCCTGGCATTCCTGG 4000

Qy        4889 AGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGT 4948
               || |||     || |||   || ||     |  |    |      | || || |||||
Db        4001 AGGAGCGGCTGATCCCGAACTTGCGCGCCAACCTGACCAGGACGCGCATTTTCACGCCCG 4060

Qy        4949 TTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTT 5008
                 |||||   |      ||| | ||||||||||||   || |||||| | ||||||||
Db        4061 CCGATTTCGCCAGTGAACTGAACGCCCATCACGGCAGCGCTTTTTCGGTCGAGCCGATCC 4120

Qy        5009 TGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATC 5068
               ||||||||  ||| |||||||| ||||| ||||| ||       ||| ||||  ||||||
Db        4121 TGACGCAATCCGCGTGGTTCCGGCCGCACAACCGGGACAAGACGATCCGCAACTTCTATC 4180

Qy        5069 TGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGG 5128
               |||| || || || || ||||| ||||||||  | || |||||  | || ||||||||||
Db        4181 TGGTCGGCGCGGGCACCCATCCGGGCGCGGGGATTCCGGGCGTCGTGGGCTCGGCCAAGG 4240

Qy        5129 CCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCA 5188
               |||| |||  | ||||||||   || |  ||      |   |  | |   | ||   |
Db        4241 CCACGGCCCAGGTGATGCTGTCCGACCTGGCGAGCGCATGAGCGATCTGGTCCT---GAC 4297

Qy        5189 AGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGA 5248
                   ||  |  | ||  |    || ||| | || || ||||| || ||||||||| |
Db        4298 CTCGACCGAGGCGATCACCCAAGGGTCGCAAAGCTTTGCCACGGCGGCCAAGCTGATGCC 4357

Qy        5249 TGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGT 5308
                 |    |  |||  |  |||| ||||||| ||||| ||||| || |||   ||||| ||
Db        4358 GCCGGGTATCCGCGACGACACGGTGATGCTCTATGCCTGGTGCCGCCACGCGGATGACGT 4417

Qy        5309 GATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGC 5368
               ||| || || ||  | |||      ||| ||   |   | || ||  | ||| |||||||
Db        4418 GATCGACGGTCAGGCCCTG------GGCAGCCGCCCCGAGGCGGTGAACGACCCGCAGGC 4471

Qy        5369 ACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGA 5428
                ||  || |    ||||     || || |     ||||     ||||       ||
Db        4472 GCGGCTGGACGGCCTGCGTGCCGACACGCTGGCGGCCCTGCAGGGCGACGGTCCGGTGAC 4531

Qy        5429 ACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGC 5488
                ||| | |||   ||| | |  |  ||||||   |  ||  |  | |||||      | |
Db        4532 GCCGCCCTTTGCCGCGCTGCGCGCGGTGGCGCGGAGGCATGATTTCCCGCAGGCCTGGCC 4591

Qy        5489 GTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGA 5548
                 | || |   | |||||||| || || ||||            | ||    |   | ||
Db        4592 CATGGACCTGATCGAAGGCTTTGCGATTGATGTCGAGGCGCGCGACTATCGCACGCTGGA 4651

Qy        5549 TGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGT 5608
               ||||  ||||   || | |||||||||||||||| |  ||||  |||||||||| |||||
Db        4652 TGACGTGCTGGAATATTCCTATCACGTCGCGGGCATCCTCGGCGTGATGATGGCCCGCGT 4711

Qy        5609 AATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCA 5668
                ||||||||||||||| |  | || || || | |||||||||  | || |||||||||||
Db        4712 GATGGGCGTGCGCGACCATCCTGTCCTGGACCGCGCCTGCGACCTGGGGCTGGCGTTCCA 4771

Qy        5669 GCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCC 5728
               ||| ||  |||| ||||||||| |  | ||  ||||    |  || || |||||||||||
Db        4772 GCTGACCCACATCGCGCGCGACGTGATCGACTATGCGCGCATCGGGCGGTGCTATCTGCC 4831

Qy        5729 GCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCG 5788
               |     |||||| || || |||||    || | |||   || |    |  | |
Db        4832 GGGGGACTGGCTGGACCACGCGGGCGCGCGGGTCGACGGGCCGGTGCCCTCAC------- 4884

Qy        5789 TGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTC 5848
                 | |  ||   | |  ||     |||  |  |||  |  || ||||||||||| |  ||
Db        4885 --CGGAACTGTACACCGTGATCCTGCGGCTGCTGGATGCCGCCGAACCCTATTACCCGTC 4942

Qy        5849 GGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCG 5908
                |||||    ||| |  ||| | | ||||  ||||  || ||| ||||| |  | ||
Db        4943 CGCGCGGGTGGGTCTGGCGGATCTGCCGCCTCGCTGCGCCTGGTCCATCCCCGCCGCGTT 5002

Qy        5909 CGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTC 5968
                 |  | |||||||  || ||  |     | |   | | |||   | | ||||
Db        5003 TCGGATCTATCGCGCCATCGGGCTGCCCATCCGCAAGGGCGGGCCGGAGGCCTATCGCCA 5062

Qy        5969 ACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGG 6018
               ||||    ||||    || |  |  ||| | |  |||||||  |  ||||
Db        5063 ACGGATCAGCACGTCCAAGGCCGCCAAAATCGGTCTGCTGGGCATCGGGG 5112
```

```
RESULT 13
ATUCRTWA
LOCUS       ATUCRTWA                5373 bp    DNA     linear   BCT 28-AUG-2002
DEFINITION  Paracoccus sp. MBIC1143 crtW, crtZ, crtY, crtI, crtB genes,
            complete cds.
ACCESSION   D58420
VERSION     D58420.1  GI:1136630
KEYWORDS    .
SOURCE      Paracoccus sp. MBIC1143
  ORGANISM  Paracoccus sp. MBIC1143
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacterales;
            Rhodobacteraceae; Paracoccus.
REFERENCE   1  (bases 1 to 5373)
  AUTHORS   Misawa,N., Kajiwara,S., Kondo,K., Yokoyama,A., Satomi,Y., Saito,T.,
            Miki,W. and Ohtani,T.
  TITLE     Canthaxanthin biosynthesis by the conversion of methylene to keto
            groups in a hydrocarbon beta-carotene by a single gene
  JOURNAL   Biochem. Biophys. Res. Commun. 209 (3), 867-876 (1995)
  MEDLINE   95251715
   PUBMED   7733978
REFERENCE   2  (bases 1 to 5373)
  AUTHORS   Misawa,N., Satomi,Y., Kondo,K., Yokoyama,A., Kajiwara,S., Saito,T.,
            Ohtani,T. and Miki,W.
  TITLE     Structure and functional analysis of a marine bacterial carotenoid
            biosynthesis gene cluster and astaxanthin biosynthetic pathway
            proposed at the gene level
  JOURNAL   J. Bacteriol. 177 (22), 6575-6584 (1995)
  MEDLINE   96062243
   PUBMED   7592436
REFERENCE   3  (bases 1 to 5373)
  AUTHORS   Misawa,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUN-1995) Norihiko Misawa, Marine Biotechnology
            Institute; 3-75-1, Heita, Kamaishi-shi, Iwate 026-0001, Japan
            (E-mail:norihiko.misawa@mbio.jp, Tel:81-193-26-6581,
            Fax:81-193-26-6584)
FEATURES             Location/Qualifiers
     source          1..5373
                     /organism="Paracoccus sp. MBIC1143"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:81397"
                     /note="Agrobacterium aurantiacum is distributed from
                     Marine Biotechnology Institute as Paracoccus sp. MBIC1143.
                     Its 16S rRNA sequence is shown in accession
                     AB008114.~cloning vector: pAK32"
     gene            139. .867
                     /gene="crtW"
     CDS             139. .867
                     /gene="crtW"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="beta-Carotene ketolase"
                     /protein_id="BAA09591.1"
                     /db_xref="GI:1136631"
                     /translation="MSAHALPKADLTATSLIVSGGIIAAWLALHVHALWFLDAAAHPI
                     LAIANFLGLTWLSVGLFIIAHDAMHGSVVPGRPRANAAMGQLVLWLYAGFSWRKMIVK
                     HMAHHRHAGTDDDPDFDHGGPVRWYARFIGTYFGWREGLLLPVIVTVYALILGDRWMY
                     VVFWPLPSILASIQLFVFGTWLPHRPGHDAFPDRHNARSSRISDPVSLLTCFHFGGYH
                     HEHHLHPTVPWWRLPSTRTKGDTA"
     gene            864. .1352
                     /gene="crtZ"
     CDS             864. .1352
                     /gene="crtZ"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="beta-Carotene hydroxylase"
                     /protein_id="BAA09592.1"
                     /db_xref="GI:1136632"
                     /translation="MTNFLIVVATVLVMELTAYSVHRWIMHGPLGWGWHKSHHEEHDH
                     ALEKNDLYGLVFAVIATVLFTVGWIWAPVLWWIALGMTVYGLIYFVLHDGLVHQRWPF
                     RYIPRKGYARRLYQAHRLHHAVEGRDHCVSFGFIYAPPVDKLKQDLKMSGVLRAEAQE
                     RT"
     gene            1349. .2509
                     /gene="crtY"
     CDS             1349. .2509
                     /gene="crtY"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="Lycopene cyclase"
                     /protein_id="BAA09593.1"
                     /db_xref="GI:1136633"
                     /translation="MTHDVLLAGAGLANGLIALALRAARPDLRVLLLDHAAGPSDGHT
                     WSCHDPDLSPDWLARLKPLRRANWPDQEVRFPRHARRLATGYGSLDGAALADAVVRSG
                     AEIRWDSDIALLDAQGATLSCGTRIEAGAVLDGRGAQPSRHLTVGFQKFVGVEIETDR
                     PHGVPRPMIMDATVTQQDGYRFIYLLPFSPTRILIEDTRYSDGGDLDDDALAAASHDY
                     ARQQGWTGAEVRRERGILPIALAHDAAGFWADHAAGPVPVGLRAGFFHPVTGYSLPYA
                     AQVADVVAGLSGPPGTDALRGAIRDYAIDRARRDRFLRLLNRMLFRGCAPDRRYTLLQ
                     RFYRMPHGLIERFYAGRLSVADQLRIVTGKPPIPLGTAIRCLPERPLLKENA"
     gene            2506. .4011
                     /gene="crtI"
     CDS             2506. .4011
                     /gene="crtI"
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="Phytoene desaturase"
                     /protein_id="BAA09594.1"
                     /db_xref="GI:1136634"
                     /translation="MNAHSPAAKTAIVIGAGFGGLALAIRLQSAGIATTLVEARDKPG
                     GRAYVWHDQGHLFDAGPTVITDPDALKELWALTGQDMARDVTLMPVSPFYRLMWPGGK
                     VFDYVNEADQLERQIAQFNPDDLEGYRRFRDYAEEVYQEGYVKLGTVPFLKLGQMLKA
                     APALMKLEAYKSVHAKVATFIKDPYLRQAFSYHTLLVGGNPFSTSSIYALNHALERRG
                     GVWFAKGGTNQLVAGMVALFERLGGQMLLNAKVARIDTDGPRATGVTLADGRALTADM
                     VASNGDVMHNYRDLLGHTARGQSRAKSLNAKRWSMSLFVLHFGLREAPKDVAHHTILF
                     GPRYKELVNEIFKGPKLAEDFSLYLHSPCTTDPEMAPPGMSTHYVLAPVPHLGRADID
                     WAVEGPRYADRILASLEERLIPNLRANLTTTRIFTPSDFASELNAHHGSAFSVEPILT
                     QSAWFRPHNRDKTIRNFYLVGAGTHPGAGIPGVVGSAKATAQVMLSDLASA"
     gene            4008. .4913
                     /gene="crtB"
     CDS             4008. .4913
                     /gene="crtB"
```

```
                     /citation=[2]
                     /codon_start=1
                     /transl_table=11
                     /product="Phytoene synthase"
                     /protein_id="BAA09595.1"
                     /db_xref="GI:1136635"
                     /translation="MSDLVLTSTEAITQGSQSFATAAKLMPPGIRDDTVMLYAWCRHA
                     DDVIDGQALGSCPEAVNDPQARLDGLRADTLAALQGDGPVTPPFACAVARRHDFPQAW
                     PMDLIEGFAMDVEARDYRTLDDVLEYSYHVAGIVGVMMARVMGVRDDPVLDRACDLGL
                     AFQLTNIARDVIDDARIGRCYLPGDWLDQAGARVDGPVPSPELYTVILRLLDAAELYY
                     ASARVGLADLPPRCAWSIAAALRIYRAIGLRIRKGGPEAYRQRISTSKAAKIGLLGIG
                     GWDVADHACRGRRVATGLWTRPHHA"
ORIGIN

  Query Match             14.8%;  Score 1037.2;  DB 1;  Length 5373;
  Best Local Similarity   57.5%;  Pred. No. 1.6e-207;
  Matches 2065;  Conservative    0;  Mismatches 1468;  Indels   57;  Gaps    9;

Qy       2456 GCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGC 2515
              ||   | | |||||   |  ||||||  |||     |  |   ||  ||| ||| |||||
Db       1277 GCCCCCGGTCGACAAGCTGAAGCAGGACCTGAAGATGTCGGGCGTGCTGCGGGCCGAGGC 1336

Qy       2516 GACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTG 2575
              |       || |  |    | || ||| |  |||  || || || || || || ||  ||
Db       1337 GCAGGAGCGCACGTGACCCATGACGTGCTGCTGGCAGGGGCGGGCCTTGCCAACGGGCTG 1396

Qy       2576 ATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGAT 2635
              || || |||   |||||      || |||  | |||      ||| |||||||    |  
Db       1397 ATCGCCCTGGCGCTGCGCGCGGCGCGGCCCGACCTGCGCGTGCTGCTGCTGGACCATGCC 1456

Qy       2636 GCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAA 2695
              ||    ||| ||| |   || |||||||| |  ||  ||  ||| ||     | ||    
Db       1457 GCAGGACCGTCAGACGGCCACACCTGGTCCTGCCACGACCCCGACCTGTCGCCGGACTGG 1516

Qy       2696 CTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCT 2755
              ||     |||||| | || |||     ||     ||| | |   |  |||||||| ||||| 
Db       1517 CTGGCGCGGCTGAAGCCCCTGCGCCGCGCCAACTGGCCCGACCAGGAGGTGCGCTTTCCC 1576

Qy       2756 GCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGC 2815
                 |    |||    |||| |   | |||  | ||  | |     |  |   | ||    
Db       1577 CGCCATGCCCGGCGGCTGGCCACCGGTTACGGGTCGCTGGACGGGGCGGCGCTGGC---G 1633

Qy       2816 CATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAA 2875
               ||      |    |  ||| | |||  |  |     |||  | | |    | | |  | 
Db       1634 GATGCGGTGGTCCGGTCGGGCGCCGAGATCCGCTGGGACAGCGACATCGCCCTGCTGGAT 1693

Qy       2876 CCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGT 2935
               |   |  || ||||||| |    |||       | |  |||   | |||  | || || 
Db       1694 GCGCAGGGGGCGACGCTGTCCTGCGGCACCCGGATCGAGGCGGGCGCGGTCCTGGACGGG 1753

Qy       2936 CGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAG 2995
              || |||  |||||| |||  ||||||    ||||||  |||   ||  |  ||    |||
Db       1754 CGGGGCGCGCAGCCGTCGCGGCATCTGACCGTGGGTTTCCAGAAATTCGTGGGTGTCGAG 1813

Qy       2996 TGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGAT 3055
                  ||   |  |  || |||||| |||  |  |||||  | ||||| ||  ||||||   
Db       1814 ATCGAGACCGACCGCCCCCACGGCGTGCCCCGCCCGATGATCATGGACGCGACCGTCACC 1873

Qy       3056 CAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATT 3115
              ||||| |  || || || ||  ||||   ||||||  ||    |    ||  |  |||| 
Db       1874 CAGCAGGACGGGTACCGCTTCATCTATCTGCTGCCCTTCTCTCCGACGCGCATCCTGATC 1933

Qy       3116 GAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATC 3175
              || || || |  ||     |      ||  ||||  ||  || |         |     |
Db       1934 GAGGACACGCGCTATTCCGATGGCGGCGATCTGGACGACGACGCGCTGGCGGCGGCGTCC 1993

Qy       3176 GCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGC 3235
                ||||||||||   ||||| ||||||||    |     | | | ||| |||   | ||||
Db       1994 CACGACTATGCCCGCCAGCAGGGCTGGACCGGGGCCGAGGTCCGGCGCGAA---CGCGGC 2050

Qy       3236 ATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGG---CAACAGCAGCGCGGC 3292
              ||  | || ||  | |||  |    |    |   | ||||||    | ||   |   || 
Db       2051 ATCCTTCCCATCGCGCTGGCCCATGATGCGGCGGGCTTCTGGGCCGATCACGCGGCGGGG 2110
```

```
Qy     3293 CAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCG 3352
            |  |    |   || |||||||| ||| | |||||| |   |||||| || ||  ||||
Db     2111 CCTGTTCCCGTGGGACTGCGCGCGGGGTTCTTTCATCCGGTCACCGGCTATTCGCTGCCC 2170

Qy     3353 TCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGT---TGCCCACCGATGCCCTCACG 3409
            |  || |    | | |||||  |||| || |  ||||    |||  |||   ||  | ||
Db     2171 TATGCGGCACAGGTGGCGGACGTGGTGGCGGGTCTGTCCGGGCCGCCCGGCACCGACGCG 2230

Qy     3410 CTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGT 3469
            ||  ||      |||      |  ||     | | |  ||||    | || ||| | ||
Db     2231 CTGCGCGGCGCCATCCGCGATTACGCGATCGACCGGGCGCGCCGCGACCGCTTTCTGCGC 2290

Qy     3470 CTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAA 3529
            ||  | ||||| ||||||||    | | |   |||  | | |||||    | || ||||
Db     2291 CTTTTGAACCGGATGCTGTTCCGCGGCTGCGCGCCCGACCGGCGCTATACCCTGCTGCAG 2350

Qy     3530 CGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTG 3589
            || || |||||  |       ||  | ||    || || || ||||| |  ||| || ||
Db     2351 CGGTTCTACCGCATGCCGCATGGACTGATCGAACGGTTCTATGCCGGCCGGCTGAGCGTG 2410

Qy     3590 CGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGC 3649
               ||| |   ||| ||  ||  ||||||||| ||  | ||| | ||    ||  | |||
Db     2411 GCGGATCAGCTGCGCATCGTGACCGGCAAGCCTCCCATTCCCCTTGGCACGGCCATCCGC 2470

Qy     3650 GCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACTT------------ 3697
               |||    |   |  |   |    | | ||   |||||   |  ||
Db     2471 TGCCTGCCCGAACGTCCCCTGCTGAAGGAAAACGCATGAACGCCCATTCGCCCGCGGCCA 2530

Qy     3698 ---------ATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAG 3748
                       ||||| ||||||||||||||||| ||||| ||||| || ||||||||
Db     2531 AGACCGCCATCGTGATCGGCGCAGGCTTTGGCGGGCTGGCCCTGGCCATCCGCCTGCAGT 2590

Qy     3749 CGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATG 3808
            | ||||||||  | |||||| |  |||||   || ||||| || ||||| ||||||||||
Db     2591 CCGCGGGCATCGCCACCACCCTGGTCGAGGCCCGGGACAAGCCCGGCGGGCGCGCCTATG 2650

Qy     3809 TGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCG 3868
            | |   | ||    |||  |  |||||| || || || || || |||||||| |||   |
Db     2651 TCTGGCACGATCAGGGCCATCTCTTCGACGCGGGCCCGACCGTCATCACCGACCCCGATG 2710

Qy     3869 CCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGC 3928
            |  |  ||||| |||   | ||| | ||  |     |        ||||   ||||||||
Db     2711 CGCTGAAAGAGCTGTGGGCCCTGACCGGGCAGGACATGGCGCGCGACGTGACGCTGATGC 2770

Qy     3929 CGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATA 3988
            ||||  ||||||||||||| |||   |||   |  || ||     | |||||||  ||
Db     2771 CGGTCTCGCCCTTCTATCGGCTGATGTGGCCGGGCGGGAAGGTCTTCGATTACGTGAACG 2830

Qy     3989 ATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATC 4048
            |       | |||||| |  |||||||||  |||||| ||| | ||  | || || || |
Db     2831 AGGCCGATCAGCTGGAACGCCAGATCGCCCAGTTCAACCCGGACGACCTGGAGGGATACC 2890

Qy     4049 GTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGG 4108
            | |  || | ||  ||  |    || || |     ||||| ||  | || || ||||| |
Db     2891 GCCGCTTCCGTGATTACGCTGAGGAGGTGTACCAGGAGGGCTACGTCAAGCTGGGCACCG 2950

Qy     4109 TGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAG 4168
            |||| || ||  || || |  | |||||    | ||||||   | ||       || | |
Db     2951 TGCCCTTCCTCAAGCTGGGCCAGATGCTCAAGGCCGCGCCCGCGCTGATGAAGTTGGAGG 3010

Qy     4169 CATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGG 4228
            | |        ||| |   || ||| ||||  || ||  |||||    ||||||| ||||
Db     3011 CCTACAAGTCGGTCCATGCCAAGGTCGCGACCTTCATCAAGGACCCCTATCTGCGGCAGG 3070

Qy     4229 CGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATA 4288
            ||||||| |  ||| |  ||||||||||||| ||||| ||  | ||    |||||||||
Db     3071 CGTTTTCGTATCACACGCTGCTGGTGGGCGGGAATCCCTTCTCGACCAGCTCGATCTATG 3130

Qy     4289 CCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCG 4348
            |  | |  || |||||||||||     | || || |||||  |    |||||||||  |
Db     3131 CGCTGAACCACGCGCTGGAGCGGCGCGGCGGGGTCTGGTTCGCCAAGGGCGGCACCAACC 3190

Qy     4349 CGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATG 4408
             ||||||   ||||||||     ||||||||     | |||||  || || | || || |
Db     3191 AGCTGGTTGCGGGCATGGTCGCGCTGTTCGAACGGCTTGGCGGGCAGATGCTGCTAAACG 3250

Qy     4409 CCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCG 4468
            || | ||    | | | || ||   |||    |||   | ||||||     | |  | ||
Db     3251 CCAAGGTCGCGCGGATCGACACGGACGGGCCGCGCGCGACCGGCGTGACCCTGGCCGACG 3310

Qy     4469 GACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACA 4528
            | ||   |||    |||| |   || |||  ||| | ||||||| |||| | |||   |
Db     3311 GGCGCGCCTTGACCGCCGACATGGTCGCCAGCAACGGCGACGTGATGCACAACTATCGCG 3370

Qy     4529 AACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCA 4588
            | |||||  ||||     | |  |    ||  || |||| |||||||||    ||||||
Db     3371 ACCTGCTGGGCCATACCGCCCGCGGGCAGAGCCGGGCGAAATCGCTGAACGCGAAGCGCT 3430

Qy     4589 TGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGC 4648
             |  ||  ||||| || || ||  |||| ||||||    || |||   |  |  | ||||
Db     3431 GGTCCATGTCGCTCTTCGTGCTGCATTTCGGCCTGCGCGAGGCGCCCAAGGACGTGGCGC 3490

Qy     4649 ACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCA 4708
            | |||||| ||   || ||||| || ||    ||| || || | ||||| |||||  ||
Db     3491 ATCACACCATCCTGTTCGGCCCCCGCTACAAGGAGCTGGTCAACGAGATCTTCAAGGGCC 3550

Qy     4709 GCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGC 4768
               ||||||| || ||||| || || || |||||  |||||||||  | ||| |||  |
Db     3551 CGAAGCTGGCCGAGGATTTCTCGCTCTATCTGCATTCGCCCTGCACGACCGACCCGGAGA 3610

Qy     4769 TGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCG 4828
            |||| || || |||     ||    ||| ||  | || |||||||||||||| |||  ||
Db     3611 TGGCGCCTCCGGGCATGTCCACGCATTACGTCCTGGCCCCGGTGCCGCATCTGGGCCGCG 3670

Qy     4829 CTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGG 4888
            | ||||| || |||      || || |||||||    ||| || ||  | ||    ||||
Db     3671 CGGACATTGATTGGGCGGTCGAGGGGCCGCGCTATGCCGACCGCATCCTGGCCAGCCTGG 3730

Qy     4889 AGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGT 4948
            || |||     || |||    | ||     |  |    ||     | || || || ||||
Db     3731 AGGAGCGGCTGATCCCGAACCTGCGCGCCAACCTGACCACGACGCGCATCTTCACCCCGT 3790

Qy     4949 TTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTT 5008
              |||||   |  |   ||| | |||||||| |||   || || ||| | ||||||||
Db     3791 CCGATTTCGCCAGCGAACTGAACGCCCATCATGGCAGCGCCTTCTCGGTCGAGCCGATCC 3850

Qy     5009 TGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATC 5068
            ||||||||  ||| |||||||| ||||| ||||||||       ||| ||||  ||||||
Db     3851 TGACGCAATCCGCGTGGTTCCGGCCGCACAACCGCGACAAGACGATCCGCAACTTCTATC 3910

Qy     5069 TGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGG 5128
            |||| || || || || ||||| ||||||||  | || |||||  | || ||||||||||
Db     3911 TGGTCGGCGCGGGCACCCATCCGGGCGCGGGGATTCCGGGCGTCGTGGGCTCGGCCAAGG 3970

Qy     5129 CCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCA 5188
            |||| |||  | ||||||||   || |  ||      |   |  | |   | ||   |
Db     3971 CCACGGCCCAGGTGATGCTGTCCGACCTGGCGAGCGCATGAGCGATCTGGTCCT---GAC 4027

Qy     5189 AGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGA 5248
                ||  |  | ||  |    || ||| | || || ||||| || ||||||||| |
Db     4028 CTCGACCGAGGCGATCACCCAAGGGTCGCAAAGCTTTGCCACGGCGGCCAAGCTGATGCC 4087

Qy     5249 TGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGT 5308
              |    |  |||  |  |||| ||||||| ||||| ||||| || |||   ||||| ||
Db     4088 GCCGGGCATCCGCGACGACACGGTGATGCTCTATGCCTGGTGCCGCCACGCGGATGACGT 4147

Qy     5309 GATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGC 5368
            ||| || || ||  | |||      ||| ||   |   | || ||  | ||| |||||||
Db     4148 GATCGACGGTCAGGCCCTG------GGCAGCTGCCCCGAGGCGGTGAACGACCCGCAGGC 4201

Qy     5369 ACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGA 5428
             ||  || |    ||||     || || |    ||| |     ||||       ||
Db     4202 GCGGCTGGACGGCCTGCGCGCCGACACGCTGGCCGCGTTGCAGGGCGACGGTCCGGTGAC 4261

Qy     5429 ACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGC 5488
```

```
                ||| | |||            |  ||||||      ||  |  | |||||      | |
Db        4262 CCCGCCCTTTGC------CTGCGCGGTGGCGCGGCGGCATGATTTTCCGCAGGCCTGGCC 4315

Qy        5489 GTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGA 5548
                 | || |   | |||||||| || |||||||            | ||    |   | ||
Db        4316 CATGGACCTGATCGAAGGCTTTGCGATGGATGTCGAGGCGCGCGACTATCGCACGCTGGA 4375

Qy        5549 TGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGT 5608
                |||  ||||   || | |||||||||||||||| | |||||  |||||||||| || ||
Db        4376 CGACGTGCTGGAATATTCCTATCACGTCGCGGGCATCGTCGGCGTGATGATGGCCCGTGT 4435

Qy        5609 AATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCA 5668
                |||||||||||||||||  | || || || |  ||||||||  | || |||||||||||
Db        4436 GATGGGCGTGCGCGACGATCCTGTCCTGGACCGTGCCTGCGACCTGGGGCTGGCGTTCCA 4495

Qy        5669 GCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCC 5728
               ||| || ||||| ||||||||| |  | || || ||    |  || || |||||||||||
Db        4496 GCTGACCAACATCGCGCGCGACGTGATCGACGACGCGCGCATCGGGCGGTGCTATCTGCC 4555

Qy        5729 GCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCG 5788
                     |||||| |||||||||||    || | |||   || |    |  |||
Db        4556 AGGCGACTGGCTGGATCAGGCGGGCGCGCGGGTCGACGGGCCGGTGCCGTCGC------- 4608

Qy        5789 TGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTC 5848
                 | | |||   | |  ||     |||  |  |||  | ||| |||| ||||||    ||
Db        4609 --CGGAGCTGTACACCGTGATCCTGCGCCTGCTGGATGCGGCTGAACTCTATTACGCGTC 4666

Qy        5849 GGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCG 5908
               ||||||    ||| |  ||| | | |||| |||||  || ||| |||||||  | ||
Db        4667 GGCGCGGGTGGGTCTGGCGGATCTGCCGCCGCGCTGCGCCTGGTCCATCGCCGCCGCGTT 4726

Qy        5909 CGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTC 5968
                 |  | |||||||  || ||  |     | |   | | |||   | | ||||
Db        4727 GCGGATCTATCGCGCCATCGGGCTGCGCATCCGCAAGGGCGGGCCGGAGGCCTATCGCCA 4786

Qy        5969 ACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGG 6018
                |||    ||||    || |  |  ||  | | ||||||||  |  ||||
Db        4787 GCGGATCAGCACGTCCAAGGCCGCCAAGATCGGGCTGCTGGGCATCGGGG 4836


RESULT 14
FAU62808/c
LOCUS       FAU62808                5188 bp    DNA     linear   BCT 18-FEB-1997
DEFINITION  Flavobacterium ATCC21588 geranylgeranyl synthase (crtE), phytoene
            synthase (crtB), phytoene desaturase (crtI), lycopene cyclase
            (crtY), b-carotene hydroxylase (crtZ) genes, complete cds.
ACCESSION   U62808
VERSION     U62808.1  GI:1842241
KEYWORDS    .
SOURCE      Flavobacterium sp. ATCC 21588
  ORGANISM  Flavobacterium sp. ATCC 21588
            Bacteria; Bacteroidetes; Flavobacteria; Flavobacteriales;
            Flavobacteriaceae; Flavobacterium.
REFERENCE   1  (bases 1 to 5188)
  AUTHORS   Pasamontes,L., Hug,D., Tessier,M., Hohmann,H.P., Schierle,J. and
            van Loon,A.P.
  TITLE     Isolation and characterization of the carotenoid biosynthesis genes
            of Flavobacterium sp. strain R1534
  JOURNAL   Gene 185 (1), 35-41 (1997)
  MEDLINE   97186694
   PUBMED   9034310
REFERENCE   2  (bases 1 to 5188)
  AUTHORS   Pasamontes,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-JUL-1996) VFCB, F. Hoffmann-La Roche Ltd., Basel,
            4070, Switzerland
FEATURES             Location/Qualifiers
     source          1. .5188
                     /organism="Flavobacterium sp. ATCC 21588"
                     /mol_type="genomic DNA"
                     /strain="R1534"
                     /db_xref="taxon:50286"
     gene            95. .982
                     /gene="crtE"
     CDS             95. .982
                     /gene="crtE"
                     /codon_start=1
                     /transl_table=11
                     /product="geranylgeranyl synthase"
                     /protein_id="AAC44848.1"
                     /db_xref="GI:1842242"
                     /translation="MTPKQQFPLRDLVEIRLAQISGQFGVVSAPLGAAMSDAALSPGK
                     RFRAVLMLMVAESSGGVCDAMVDAACAVEMVHAASLIFDDMPCMDDARTRRGQPATHV
                     AHGEGRAVLAGIALITEAMRILGEARGATPDQRARLVASMSRAMGPVGLCAGQDLDLH
                     APKDAAGIEREQDLKTGVLFVAGLEMLSIIKGLDKAETEQLMAFGRQLGRVFQSYDDL
                     LDVIGDKASTGKDTARDTAAPGPKGGLMAVGQMGDVAQHYRASRAQLDELMRTRLFRG
                     GQIADLLARVLPHDIRRSA"
     gene            complement(979. .1890)
                     /gene="crtB"
     CDS             complement(979. .1890)
                     /gene="crtB"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene synthase"
                     /protein_id="AAC44849.1"
                     /db_xref="GI:1842243"
                     /translation="MTDLTATSEAAIAQGSQSFAQAAKLMPPGIREDTVMLYAWCRHA
                     DDVIDGQVMGSAPEAGGDPQARLGALRADTLAALHEDGPMSPPFAALRQVARRHDFPD
                     LWPMDLIEGFAMDVADREYRSLDDVLEYSYHVAGVVGVMMARVMGVQDDAVLDRACDL
                     GLAFQLTNIARDVIDDAAIGRCYLPADWLAEAGATVEGPVPSDALYSVIIRLLDAAEP
                     YYASARQGLPHLPPRCAWSIAAALRIYRAIGTRIRQGGPEAYRQRISTSKAAKIGLLA
                     RGGLDAAASRLRGGEISRDGLWTRPRA"
     gene            complement(1887. .3371)
                     /gene="crtI"
     CDS             complement(1887. .3371)
                     /gene="crtI"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene desaturase"
                     /protein_id="AAC44850.1"
                     /db_xref="GI:1842244"
                     /translation="MSSAIVIGAGFGGLALAIRLQSAGIATTIVEARDKPGGRAYVWN
                     DQGHVFDAGPTVVTDPDSLRELWALSGQPMERDVTLLPVSPFYRLTWADGRSFEYVND
                     DDELIRQVASFNPADVDGYRRFHDYAEEVYREGYLKLGTTPFLKLGQMLNAAPALMRL
                     QAYRSVHSMVARFIQDPHLRQAFSFHTLLVGGNPFSTSSIYALIHALERRGGVWFAKG
                     GTNQLVAGMVALFERLGGTLLLNARVTRIDTEGDRATGVTLLDGRQLRADTVASNGDV
                     MHSYRDLLGHTRRGRTKAAILNRQRWSMSLFVLHFGLSKRPENLAHHSVIFGPRYKGL
                     VNEIFNGPRLPDDFSMYLHSPCVTDPSLAPEGMSTHYVLAPVPHLGRADVDWEAEAPG
                     YAERIFEELERRAIPDLRKHLTVSRIFSPADFSTELSAHHGSAFSVEPILTQSAWFRP
                     HNRDRAIPNFYIVGAGTHPGAGIPGVVGSAKATAQVMLSDLAVA"
     gene            complement(3368. .4516)
                     /gene="crtY"
     CDS             complement(3368. .4516)
                     /gene="crtY"
                     /codon_start=1
                     /transl_table=11
                     /product="lycopene cyclase"
                     /protein_id="AAC44851.1"
                     /db_xref="GI:1842245"
                     /translation="MSHDLLIAGAGLSGALIALAVRDRRPDARIVMLDARSGPSDQHT
                     WSCHDTDLSPEWLARLSPIRRGEWTDQEVAFPDHSRRLTTGYGSIEAGALIGLLQGVD
                     LRWNTHVATLDDTGATLTDGSRIEAACVIDARGAVETPHLTVGFQKFVGVEIETDAPH
                     GVERPMIMDATVPQMDGYRFIYLLPFSPTRILIEDTRYSDGGDLDDGALAQASLDYAA
                     RRGWTGQEMRRERGILPIALAHDAIGFWRDHAQGAVPVGLGAGLFHPVTGYSLPYAAQ
                     VADAIAARDLTTASARRAVRGWAIDRADRDRFLRLLNRMLFRGCPPDRRYRLLQRFYR
                     LPQPLIERFYAGRLTLADRLRIVTGRPPIPLSQAVRCLPERPLLQERA"
     gene            complement(4513. .5022)
                     /gene="crtZ"
     CDS             complement(4513. .5022)
                     /gene="crtZ"
                     /codon_start=1
                     /transl_table=11
                     /product="b-carotene hydroxylase"
                     /protein_id="AAC44852.1"
                     /db_xref="GI:1842246"
                     /translation="MSTWAAILTVILTVAAMELTAYSVHRWIMHGPLGWGWHKSHHDE
```

```
                    DHDHALEKNDLYGVIFAVISIVLFAIGAMGSDLAWWLAVGVTCYGLIYYFLHDGLVHG
                    RWPFRYVPKRGYLRRVYQAHRMHHAVHGRENCVSFGFIWAPSVDSLKAELKRSGALLK
                    DREGADRNT"
ORIGIN

  Query Match             14.8%;  Score 1035.4;  DB 1;  Length 5188;
  Best Local Similarity   56.5%;  Pred. No. 3.8e-207;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy       2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545

Db       4558 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 4499

Qy       2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605

Db       4498 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 4439

Qy       2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665

Db       4438 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 4379

Qy       2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725

Db       4378 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 4319

Qy       2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785

Db       4318 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 4259

Qy       2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845

Db       4258 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 4208

Qy       2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905

Db       4207 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 4148

Qy       2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965

Db       4147 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 4088

Qy       2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025

Db       4087 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCATGGCGTCGAG 4028

Qy       3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085

Db       4027 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 3968

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145

Db       3967 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 3908

Qy       3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205

Db       3907 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 3848

Qy       3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265

Db       3847 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 3788

Qy       3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325

Db       3787 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 3728

Qy       3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385

Db       3727 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 3668

Qy       3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445

Db       3667 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 3608

Qy       3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505

Db       3607 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 3551

Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565

Db       3550 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 3491

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625

Db       3490 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 3431

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685

Db       3430 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 3371

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745

Db       3370 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 3311

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805

Db       3310 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 3251

Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865

Db       3250 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 3191

Qy       3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925

Db       3190 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 3131

Qy       3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985

Db       3130 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 3071

Qy       3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045

Db       3070 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 3011

Qy       4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105

Db       3010 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 2951

Qy       4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165

Db       2950 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 2891

Qy       4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225

Db       2890 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 2831

Qy       4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285

Db       2830 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 2771

Qy       4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345

Db       2770 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 2711

Qy       4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405

Db       2710 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 2651

Qy       4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465

Db       2650 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 2591

Qy       4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525

Db       2590 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 2531

Qy       4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
```

```
Db       2530 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 2471

Qy       4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645

Db       2470 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 2411

Qy       4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705

Db       2410 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 2351

Qy       4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765

Db       2350 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 2291

Qy       4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825

Db       2290 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 2231

Qy       4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885

Db       2230 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 2171

Qy       4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945

Db       2170 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 2111

Qy       4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005

Db       2110 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 2051

Qy       5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065

Db       2050 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 1991

Qy       5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125

Db       1990 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 1931

Qy       5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185

Db       1930 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 1871

Qy       5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245

Db       1870 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 1814

Qy       5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305

Db       1813 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 1754

Qy       5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365

Db       1753 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 1700

Qy       5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425

Db       1699 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 1640

Qy       5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485

Db       1639 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 1580

Qy       5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545

Db       1579 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 1520

Qy       5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605

Db       1519 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 1460

Qy       5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
```

```
Db       1459 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 1400

Qy       5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725

Db       1399 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 1340

Qy       5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785

Db       1339 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 1289

Qy       5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845

Db       1288 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 1229

Qy       5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905

Db       1228 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 1169

Qy       5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965

Db       1168 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 1109

Qy       5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025

Db       1108 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 1049

Qy       6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085

Db       1048 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 989

Qy       6086 TCG 6088

Db        988 GCG 986


RESULT 15
AR169831/c
LOCUS       AR169831                  8625 bp    DNA     linear   PAT 17-DEC-2001
DEFINITION  Sequence 1 from patent US 6291204.
ACCESSION   AR169831
VERSION     AR169831.1  GI:17907751
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 8625)
  AUTHORS   Pasamontes,L. and Tsygankov,Y.
  TITLE     Fermentative carotenoid production
  JOURNAL   Patent: US 6291204-A 1 18-SEP-2001;
FEATURES             Location/Qualifiers
     source          1. .8625
                     /organism="unknown"
                     /mol_type="unassigned DNA"
ORIGIN

  Query Match             14.8%;  Score 1035.4;  DB 6;  Length 8625;
  Best Local Similarity   56.5%;  Pred. No. 3.8e-207;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy       2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545

Db       6984 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 6925

Qy       2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605

Db       6924 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 6865

Qy       2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665

Db       6864 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 6805

Qy       2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
```

```
Db       6804 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 6745

Qy       2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785

Db       6744 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 6685

Qy       2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845

Db       6684 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 6634

Qy       2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905

Db       6633 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 6574

Qy       2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965

Db       6573 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 6514

Qy       2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025

Db       6513 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCATGGCGTCGAG 6454

Qy       3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085

Db       6453 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 6394

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145

Db       6393 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 6334

Qy       3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205

Db       6333 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6274

Qy       3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265

Db       6273 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6214

Qy       3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325

Db       6213 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6154

Qy       3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385

Db       6153 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6094

Qy       3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445

Db       6093 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGCCATCGATCGCGCG 6034

Qy       3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505

Db       6033 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 5977

Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565

Db       5976 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 5917

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625

Db       5916 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 5857

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685

Db       5856 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 5797

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745

Db       5796 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 5737

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805

Db       5736 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 5677
```

```
Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865

Db       5676 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 5617

Qy       3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925

Db       5616 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 5557

Qy       3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985

Db       5556 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 5497

Qy       3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045

Db       5496 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 5437

Qy       4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105

Db       5436 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 5377

Qy       4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165

Db       5376 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5317

Qy       4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225

Db       5316 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5257

Qy       4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285

Db       5256 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5197

Qy       4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345

Db       5196 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5137

Qy       4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405

Db       5136 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5077

Qy       4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465

Db       5076 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5017

Qy       4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525

Db       5016 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 4957

Qy       4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585

Db       4956 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 4897

Qy       4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645

Db       4896 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 4837

Qy       4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705

Db       4836 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 4777

Qy       4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765

Db       4776 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 4717

Qy       4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825

Db       4716 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 4657

Qy       4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885

Db       4656 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 4597
```

```
Qy      4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
             ||||||  | |  ||| || |   | ||  ||||  |         | | || || |  |
Db      4596 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 4537

Qy      4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
             |    |||||  ||  |   |||   ||||||||||||   || || ||| | |||||||
Db      4536 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 4477

Qy      5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
             |  ||||||||  |||||||||||||||||||||||||||   ||  |||   ||  |||
Db      4476 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 4417

Qy      5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
             |  | ||||| || || |||||||| || |||||| | || || ||  | ||    ||||
Db      4416 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 4357

Qy      5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
             ||||||| ||   | | ||||||  ||| | ||||     |      | |
Db      4356 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4297

Qy      5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
                 | | ||    |||| |||  ||| ||| | || |||||    || ||||||||| |
Db      4296 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4240

Qy      5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
                  |    |  |||      ||| | ||||| ||||| |||||  | ||    |||||
Db      4239 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4180

Qy      5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
              ||||| || |||||   | ||||    | |  |  |      || | ||  ||| | ||
Db      4179 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4126

Qy      5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
             ||| ||  ||  |   ||||     || || |    |||      ||  |    |  |
Db      4125 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4066

Qy      5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
                 ||| | ||   |||| | |   | || ||       ||  |  | ||| | |
Db      4065 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4006

Qy      5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
             | || | || |   | || || ||||| |||||||      |     |  |||   ||| |
Db      4005 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 3946

Qy      5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
              ||||||  ||||   || | |||  ||||||||||| || || ||  |||||||||||||
Db      3945 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 3886

Qy      5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
              || ||||||||||  ||||| |||||||| |||| |||||||||| | || || |||||
Db      3885 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 3826

Qy      5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
             |||||| || ||||| || |||||| |  | || |||||||  |  || |||||||||||
Db      3825 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 3766

Qy      5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
             |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db      3765 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 3715

Qy      5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
             |   |  ||||||   ||  |       ||  |  | |  | ||| || |||||||||
Db      3714 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 3655

Qy      5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
             |||||||||    ||  | |||  | | |||| ||||| |||||| | |||||  | ||
Db      3654 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 3595

Qy      5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
              |   |  | ||||||| ||| ||        | | ||| |  ||     | ||||
Db      3594 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 3535

Qy      5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                 |||    ||||    || | ||  ||  | | ||| |||| |   || |
Db      3534 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 3475

Qy      6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
              |||  | | |||||  ||    |    |||    | |  |  || ||||||   ||  ||
Db      3474 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 3415

Qy      6086 TCG 6088
              ||
Db      3414 GCG 3412
```

Search completed: January 25, 2005, 18:20:30
Job time : 28284 secs

```
                    GenCore version 5.1.6
             Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 24, 2005, 22:47:11 ; Search time 2935 Seconds
                                             (without alignments)
                                             12518.125 Million cell updates/sec

Title:          US-10-808-979-18
Perfect score:  6999
Sequence:       1 gcgatggcaatggaaaatgt..........gacgttttagacatgcaata 6999

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4134886 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters:      8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       N_Geneseq_23Sep04:*
                1:  geneseqn1980s:*
                2:  geneseqn1990s:*
                3:  geneseqn2000s:*
                4:  geneseqn2001as:*
                5:  geneseqn2001bs:*
                6:  geneseqn2002as:*
                7:  geneseqn2002bs:*
                8:  geneseqn2003as:*
                9:  geneseqn2003bs:*
                10:  geneseqn2003cs:*
                11:  geneseqn2003ds:*
                12:  geneseqn2004s:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                   SUMMARIES
                       %
Result                 Query
  No.     Score   Match Length DB  ID                           Description
 ---------------------------------------------------------------------------
     1     2623    37.5    6918  2  AAQ06299                    Aaq06299 Sequence
     2   2619.6    37.4    8609  12  ADQ14666                    Adq14666 Reporter
     3   2619.6    37.4    8609  12  ADQ26394                    Adq26394 Carotenoi
     4   2619.6    37.4    8609  12  ADQ77262                    Adq77262 Nucleotid
     5   2619.6    37.4    8609  12  ADQ48673                    Adq48673 Plasmid p
     6   2135.2    30.5   12753  12  ADM98599                    Adm98599 Geranylge
     7   1900.6    27.2    4624  12  ADP74122                    Adp74122 Pantoea a
     8   1618.8    23.1    7494  8  ABZ69177                    Abz69177 Vector co
     9   1618.8    23.1    8547  8  ABZ69178                    Abz69178 Vector co
c   10   1374.8    19.6 110000  10  ACF67367_41                  Continuation (42 o
    11   1374.8    19.6 110000  10  ACF65388_06                  Continuation (7 of
    12   1374.8    19.6 110000  10  ACF65388_07                  Continuation (8 of
c   13     1037    14.8    8625  2  AAV40146                    Aav40146 Flavobact
c   14     1037    14.8   11233  2  AAV40151                    Aav40151 DNA seque
c   15   1035.4    14.8    8625  2  AAT45143                    Aat45143 Flavobact
    16    827.4    11.8    1493  2  AAX90713                    Aax90713 Erwinia u
    17    826.2    11.8    1479  2  AAQ06296                    Aaq06296 Sequence
    18    826.2    11.8    1479  2  AAV84081                    Aav84081 Carotenoi
    19    826.2    11.8    1479  2  AAX19119                    Aax19119 Erwinia u
    20    826.2    11.8    1479  12  ADO05014                    Ado05014 P. ananat
    21    808.6    11.6    1479  6  AAD35512                    Aad35512 Pantoea s
    22    808.6    11.6    1479  8  ABT14193                    Abt14193 Pantoea s
    23    808.6    11.6    1479  8  ACC44762                    Acc44762 Pantoea s
    24    808.6    11.6    1479  12  ADQ14630                    Adq14630 Pantoea s
    25    808.6    11.6    1479  12  ADQ77205                    Adq77205 Nucleotid
    26    808.6    11.6    1479  12  ADQ48640                    Adq48640 Pantoea s
    27    713.4    10.2    1518  2  AAQ13719                    Aaq13719 Phytoene_
    28    713.4    10.2    1518  2  AAT40792                    Aat40792 Phytoene
    29    713.4    10.2    1518  2  AAT37094                    Aat37094 Phytoene
    30    713.4    10.2    1518  2  AAT91544                    Aat91544 Erwinia h
    31    708.6    10.1    1482  10  ACF70993                    Acf70993 Photorhab
    32    708.4    10.1    1517  2  AAQ13720                    Aaq13720 Phytoene_
    33    708.4    10.1    1522  2  AAT40793                    Aat40793 Phytoene
    34    708.4    10.1    1522  2  AAT37095                    Aat37095 Phytoene
    35    708.4    10.1    1522  2  AAT91546                    Aat91546 Erwinia h
    36    692.6     9.9    1521  10  ADC99016                    Adc99016 Bradyrhiz
    37    622.8     8.9    1497  2  AAV73181                    Aav73181 C. utilis
    38    515.6     7.4    1198  2  AAQ13718                    Aaq13718 Phytoene
    39    515.6     7.4    1198  2  AAT40791                    Aat40791 Phytoene
    40    515.6     7.4    1198  2  AAT37093                    Aat37093 E. herbic
    41    515.6     7.4    1198  2  AAT41743                    Aat41743 Phytoene
    42    515.6     7.4    1198  2  AAT91543                    Aat91543 Erwinia h
    43    461.4     6.6    1581  2  AAT31798                    Aat31798 Erythroba
    44    449.6     6.4    2034  11  ADQ96835                    Adq96835 CrtIcrtZ
    45    449.6     6.4    2051  11  ADQ96934                    Adq96934 CrtIcrtZ

                                  ALIGNMENTS

RESULT 1
AAQ06299
ID   AAQ06299 standard; DNA; 6918 BP.
XX
AC   AAQ06299;
XX
DT   24-OCT-2003  (revised)
DT   28-JAN-1991  (first entry)
XX
DE   Sequence encoding six enzymes of the carotenoid biosynthetic pathway.
XX
KW   Carotenoid biosynthesis; vitamin A; cancer; food coloring; ss.
XX
OS   Pantoea ananatis.
XX
FH   Key             Location/Qualifiers
FT   mat_peptide     225. .1130
FT                   /*tag=  a
FT                   /note= "See AAQ06293"
FT   mat_peptide     1142. .2435
FT                   /*tag=  b
FT                   /note= "See AAQ06294"
FT   mat_peptide     2421. .3567
FT                   /*tag=  c
FT                   /note= "See AAQ06295"
FT   mat_peptide     3582. .5057
FT                   /*tag=  d
FT                   /note= "See AAQ06296"
FT   mat_peptide     5096. .5983
FT                   /*tag=  e
FT                   /note= "See AAQ06297"
FT   mat_peptide     complement(5928. .6452)
FT                   /*tag=  f
FT                   /note= "See AAQ06298"
XX
PN   EP393690-A.
XX
PD   24-OCT-1990.
XX
PF   20-APR-1990;    90EP-00107493.
XX
PR   21-APR-1989;    89JP-00103078.
PR   05-MAR-1990;    90JP-00053255.
XX
```

```
PA   (KIRI ) KIRIN BEER KK.
XX
PI   Misawa N,  Kobayashi K,  Nakamura K;
XX
DR   WPI; 1990-322212/43.
XX
PT   DNA sequences encoding enzymes for carotenoid biosynthesis - for prodn.
PT   of carotenoid cpds. e.g. beta-carotene, lycopene, phytoene, etc.
XX
PS   Disclosure; Fig 7; 40pp; English.
XX
CC   Gene products are useful for the synthesis of carotenoids, useful as food
CC   coloring, vitamin A precursor, and possibly in prevention of cancer. See
CC   also AAQ06293-9. (Updated on 24-OCT-2003 to standardise OS field)
XX
SQ   Sequence 6918 BP; 1522 A; 1876 C; 1881 G; 1639 T; 0 U; 0 Other;

  Query Match             37.5%;  Score 2623;  DB 2;  Length 6918;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 4006;  Conservative    0;  Mismatches 2170;  Indels   21;  Gaps    6;

Qy        400 CTTGAACATTTACTGCCTGCCGGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCC 459
              ||||| || ||| |||| |  |      || | |||   ||| |||||||||||||||
Db        303 CTTGATCAGTTATTGCCCGTGGAGGGAGAACGGGATGTTGTGGGTGCCGCGATGCGTGAA 362

Qy        460 GGAACGCTGGCGCAGGGCAAACGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGAT 519
              ||  ||||||| | ||| ||||||||||| ||  | || ||| ||||| | || ||||||
Db        363 GGTGCGCTGGCACCGGGAAAACGTATTCGCCCCATGTTGCTGTTGCTGACCGCCCGCGAT 422

Qy        520 ATGGGTTGCGAGCTGACGCAAAATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTG 579
               |||||||||   | |  ||  | ||  | || ||| | |||||||| |||||||||||
Db        423 CTGGGTTGCGCTGTCAGCCATGACGGATTACTGGATTTGGCCTGTGCGGTGGAAATGGTC 482

Qy        580 CACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGT 639
              |||||||| |||||||| || || || || ||||  |||||  | ||| || |||| ||
Db        483 CACGCGGCTTCGCTGATCCTTGACGATATGCCCTGCATGGACGATGCGAAGCTGCGGCGC 542

Qy        640 GGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCG 699
              || ||||||||| | |||    | |  || ||  | ||||| || || || ||  | ||
Db        543 GGACGCCCTACCATTCATTCTCATTACGGAGAGCATGTGGCAATACTGGCGGCGGTTGCC 602

Qy        700 CTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAA 759
               |||| ||    || ||||  || ||||||  ||||   ||  |  |  |  |    |||
Db        603 TTGCTGAGTAAAGCCTTTGGCGTAATTGCCGATGCAGATGGCCTCACGCCGCTGGCAAAA 662

Qy        760 TCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAA 819
                |   ||| || ||||||| ||    ||| ||||| ||||| || || || || || ||
Db        663 AATCGGGCGGTTTCTGAACTGTCAAACGCCATCGGCATGCAAGGATTGGTTCAGGGTCAG 722

Qy        820 TTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAAC 879
              ||| ||||||||   || ||     ||   |||||| | ||||| ||    ||||| ||
Db        723 TTCAAGGATCTGTCTGAAGGGGATAAGCCGCGCAGCGCTGAAGCTATTTTGATGACGAAT 782

Qy        880 GAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGAC 939
               |  | |||||||||  |||||||| | ||| |  |||| |||||  ||||||  ||  |
Db        783 CACTTTAAAACCAGCACGCTGTTTTGTGCCTCCATGCAGATGGCCTCGATTGTTGCGAAT 842

Qy        940 GCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTT 999
              || ||     | | |||  |  | ||       ||  | |  ||| | || ||||| |||
Db        843 GCCTCCAGCGAAGCGCGTGATTGCCTGCATCGTTTTTCACTTGATCTTGGTCAGGCATTT 902

Qy       1000 CAACTGCTCGACGACCTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGAT 1059
              |||||||| |||||  |  |||| ||    |   |||||||||| |||    | |||||
Db        903 CAACTGCTGGACGATTTGACCGATGGCATGACCGACACCGGTAAGGATAGCAATCAGGAC 962

Qy       1060 CAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGC 1119
                 || ||||| ||||||||  |  || | ||  |    |  |  ||    || ||| |
Db        963 GCCGGTAAATCGACGCTGGTCAATCTGTTAGGCCCGAGGGCGGTTGAAGAACGTCTGAGA 1022

Qy       1120 GATCACCTGCGCAGCGCAGATGCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACT 1179
               | || || |     ||   ||  || ||  |    ||||||||| | |||   |||||||
Db       1023 CAACATCTTCAGCTTGCCAGTGAGCATCTCTCTGCGGCCTGCCAACACGGGCACGCCACT 1082

Qy       1180 CGCCAATATATGCACGCGCTGTTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTC 1239
              |  || | ||| || ||   |||| |  || | || ||     |||  | ||   |   |
Db       1083 CAACATTTTATTCAGGCCTGGTTTGACAAAAAACTCGCTGCCGTCAGTTAAGGATGCTGC 1142

Qy       1240 AGCCGGTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGC 1299
              |     || |||| || |||| |||||| |||||  | ||||||||  ||| ||| || |
Db       1143 A-----TGAGCCATTTCGCGGCGATCGCACCGCCTTTTTACAGCCATGTTCGCGCATTAC 1197

Qy       1300 AGGCGTTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCG 1359
              ||    | || ||    ||| | ||||||||| |||||  | || || || ||||||  ||
Db       1198 AGAATCTCGCTCAGGAACTGGTCGCGCGCGGTCATCGGGTGACCTTTATTCAGCAATACG 1257

Qy       1360 ATGCCCGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGC 1419
              ||         ||| |      ||||  ||| |  |||  | |||||||   | | |  |
Db       1258 ATATTAAACACTTGATCGATAGCGAAACCATTGGATTTCATTCCGTCGGGACAGACAGCC 1317

Qy       1420 ATCCTGCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGC 1479
              ||||  ||||  || |  ||| |||| | || | |  |||    ||    || | |||
Db       1318 ATCCCCCCGGCGCGTTAACGCGCGTGCTACACCTGGCGGCTCATCCTCTGGGGCCGTCAA 1377

Qy       1480 TGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTG 1539
              || |     | ||| | ||  | |||  |  |||||||||||||||||||||||| ||
Db       1378 TGCTGAAGCTCATCAATGAAATGGCGCGCACCACCGATATGCTGTGCCGCGAACTCCCCC 1437

Qy       1540 CGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCG 1599
               || | | || |   ||  | ||||||||||| || |  ||  ||||||||| |||| ||||
Db       1438 AGGCATTTAACGATCTGGCCGTCGATGGCGTCATTGTTGATCAAATGGAACCGGCAGGCG 1497

Qy       1600 GATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCA 1659
                 | || ||||||||| |||   ||||||||| | || || ||||||||  ||||  |||
Db       1498 CGCTCGTTGCTGAAGCACTGGGACTGCCGTTTATCTCTGTCGCCTGCGCGCTGCCTCTCA 1557

Qy       1660 ATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAG 1719
              |||||||| | |  || || || ||||| ||||| |||   |  |        |||   |
Db       1558 ATCGTGAACCGGATATGCCCCTGGCGGTTATGCCTTTCGAATACGGGACCAGCGACGCGG 1617

Qy       1720 CGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCG 1779
              | |   |||||| |   |||||     | || |||||  |  | ||||||||||| | |
Db       1618 CTCGCGAACGTTATGCCGCCAGTGAAAAAATTTATGACTGGCTAATGCGTCGTCATGACC 1677

Qy       1780 ACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGT 1839
                || ||   | |||||   |      |    ||  |    ||||      | || ||||
Db       1678 GTGTCATTGCCGAACACAGCCACAGAATGGGCTTAGCCCCCCGGCAAAAGCTTCACCAGT 1737

Qy       1840 GCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGC 1899
              |  | ||||| ||||| |||||||||||| | || || |   | |||||||| ||  |
Db       1738 GTTTTTCGCCACTGGCGCAAATCAGCCAGCTTGTTCCTGAACTGGATTTTCCCCGCAAAG 1797

Qy       1900 AACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGC 1959
                 | || || || | ||| ||||| ||||| || ||||   ||    | ||  |  ||
Db       1798 CGTTACCGGCTTGTTTTCATGCCGTCGGGCCTCTGCGCGAAACGCACGCACCGTCAACGT 1857

Qy       1960 TCCATGCGCCCTG---GCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGC 2016
                    | |  |      || |        | ||   | |||  |||||||||||| ||||
Db       1858 CTTCATCCCGTTATTTTACATCCTCAGAAAAACCCCGGATTTTCGCCTCGCTGGGCACGC 1917

Qy       2017 TGCAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGC 2076
              | || || || || |   ||||||||       | | | | || ||     |  |
Db       1918 TTCAGGGACACCGTTATGGGCTGTTTAAAACGATAGTGAAAGCCTGTGAAGAAATTGACG 1977

Qy       2077 TATCGCTGGTGATCGCCCATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGC 2136
                  |||  || | ||||| |||||  |  | |  | |   |||    |  |||||| ||
Db       1978 GTCAGCTCCTGTTAGCCCACTGTGGTCGTCTTACGGACTCTCAGTGTGAAGAGCTGGCGC 2037

Qy       2137 TCGCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGC 2196
                    |     |   ||||  |||||| | |||||||   || || ||    || ||||
Db       2038 GAAGCCGTCATACACAGGTGGTGGATTTTGCCGATCAGTCAGCCGCGCTGTCTCAGGCGC 2097

Qy       2197 AGCTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGC 2256
              |||||   ||||| || | |||  | || |  |  ||||| ||  |  | | | | ||||
Db       2098 AGCTGGCGATCACCCACGGCGGCATGAATACGGTACTGGACGCGATTAATTACCGGACGC 2157

Qy       2257 CGATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGC 2316
```

```
Db         2158 CCCTTTTAGCGCTTCCGCTGGCCTTTGATCAGCCCGGCGTCGCGTCACGCATCGTTTATC 2217

Qy         2317 ATGACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGC 2376
Db         2218 ACGGCATCGGCAAGCGTGCTTCCCGCTTTACCACCAGCCATGCTTTGGCTCGTCAGATGC 2277

Qy         2377 AACAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGC 2436
Db         2278 GTTCATTGCTGACCAACGTCGACTTTCAGCAGCGCATGGCGAAAATCCAGACAGCCCTTC 2337

Qy         2437 AGCGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGC 2496
Db         2338 GTTTGGCAGGGGGCACCATGGCCGCTGCCGATATCATTGAGCAGGTTATGTG-CACCGGT 2396

Qy         2497 AAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGC 2556
Db         2397 CAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCGTGGGGGC 2456

Qy         2557 TGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATG 2616
Db         2457 TGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATATGCGTAT 2516

Qy         2617 CCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAG 2676
Db         2517 TTTGCTTATCGACGCCGCACCCCAGGCGGGCGGAATCATACGTGGTCATTTCACCACGA 2576

Qy         2677 CGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGG 2736
Db         2577 TGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACTGGCCCGA 2636

Qy         2737 TTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGC 2796
Db         2637 CTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTTGTATTAC 2696

Qy         2797 ATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACAC 2856
Db         2697 TTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGATGGATAC 2756

Qy         2857 AGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGC 2916
Db         2757 CGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTATCGGTGC 2816

Qy         2917 GCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCA 2976
Db         2817 CCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGGGCTTCCA 2876

Qy         2977 GGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCT 3036
Db         2877 GGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTCCCATTAT 2936

Qy         3037 GATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAG 3096
Db         2937 CATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGCCGCTCTC 2996

Qy         3097 CGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAA 3156
Db         2997 GCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAGATCCTGA 3056

Qy         3157 CACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCT 3216
Db         3057 ATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTCAGACACT 3116

Qy         3217 GCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTG 3276
Db         3117 GCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACGCATTCTG 3176

Qy         3277 GCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCAC 3336
Db         3177 G---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATCCTACCAC 3233

Qy         3337 CGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCAC 3396
Db         3234 CGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTGATGTCTT 3293

Qy         3397 CGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACA 3456
Db         3294 TACGTCGGCCTCAATTCACCATGCCATTACGCATTTGCCCGCGAGCGCTGGCAGCAGCA 3353

Qy         3457 GCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTG 3516
Db         3354 GGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATTCACGCTG 3413

Qy         3517 GCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGG 3576
Db         3414 GCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTTATGCGGG 3473

Qy         3577 GCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGG 3636
Db         3474 AAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTCCGGTATT 3533

Qy         3637 TGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACT 3696
Db         3534 AGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAAACCAACT 3593

Qy         3697 TATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGC 3756
Db         3594 ACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGCTGCGGGG 3653

Qy         3757 ATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAG 3816
Db         3654 ATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGTCTACGAG 3713

Qy         3817 GACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGAA 3876
Db         3714 GATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTCTCACCGATCCCAGTGCCATTGAA 3773

Qy         3877 GAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACG 3936
Db         3774 GAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCCGGTTACG 3833

Qy         3937 CCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCCG 3996
Db         3834 CCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGATCAAACC 3893

Qy         3997 CTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTT 4056
Db         3894 CGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCGTCAGTTT 3953

Qy         4057 CTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTT 4116
Db         3954 CTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGTCCCTTTT 4013

Qy         4117 CTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGC 4176
Db         4014 TTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGCATGGAGA 4073

Qy         4177 AGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCC 4236
Db         4074 AGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGCGTTTTCT 4133

Qy         4237 TTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAATT 4296
Db         4134 TTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATACGTTGATA 4193

Qy         4297 CATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTG 4356
Db         4194 CACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGCATTAGTT 4253

Qy         4357 CAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTG 4416
Db         4254 CAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGCCAGAGTC 4313

Qy         4417 AGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGC 4476
Db         4314 AGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGGTCGCAGG 4373
```

```
Qy       4477 TTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCTT 4536
Db       4374 TTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGACCTGTTA 4433

Qy       4537 CGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAAC 4596
Db       4434 AGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCATGAGTAAC 4493

Qy       4597 TCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACACC 4656
Db       4494 TCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCATCACACG 4553

Qy       4657 GTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTG 4716
Db       4554 GTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGATGGCCTC 4613

Qy       4717 GCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCG 4776
Db       4614 GCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACTGGCGCCT 4673

Qy       4777 CCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATC 4836
Db       4674 GAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGCGAACCTC 4733

Qy       4837 GACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCAC 4896
Db       4734 GACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGAGCAGCAT 4793

Qy       4897 TACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTTT 4956
Db       4794 TACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTTTGATTTT 4853

Qy       4957 CGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAA 5016
Db       4854 CGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCTTACCCAG 4913

Qy       5017 AGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGT 5076
Db       4914 AGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCTGGTCGGC 4973

Qy       5077 GCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCC 5136
Db       4974 GCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGCGACAGCA 5033

Qy       5137 AGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACGC 5196
Db       5034 GGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCATGCGGTCG 5090

Qy       5197 AAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGA 5256
Db       5091 AAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGATGCAAAAA 5150

Qy       5257 CGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGATG 5316
Db       5151 CCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTTATTGACG 5210

Qy       5317 GGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGC 5376
Db       5211 ATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAACGTCTGA 5270

Qy       5377 AGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCGT 5436
Db       5271 TGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAACCGGCGT 5330

Qy       5437 TTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGATC 5496
Db       5331 TTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCGTTTGATC 5390

Qy       5497 ATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACGC 5556
Db       5391 ATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGC 5450

Qy       5557 TGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCG 5616
Db       5451 TGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATCATGGGCG 5510

Qy       5617 TGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTA 5676
Db       5511 TGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAGTTGACCA 5570

Qy       5677 ACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCT 5736
Db       5571 ATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCGGCAAGCT 5630

Qy       5737 GGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCGC 5796
Db       5631 GGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGTCAGGCGC 5690

Qy       5797 TCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGAT 5856
Db       5691 TGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCTGCCACAG 5750

Qy       5857 CCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTT 5916
Db       5751 CCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAGCAGGTTT 5810

Qy       5917 ATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGC 5976
Db       5811 ACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAGCGGCAGT 5870

Qy       5977 GCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTT 6036
Db       5871 CAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCCCTTACTT 5930

Qy       6037 CGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTT 6096
Db       5931 CCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC------TCT 5984

Qy       6097 ACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAA 6156
Db       5985 AGCGCCATGTCTTTCCCGGAGCGTCGCCTGAAGTTTTGACAGGGGCGGCGCATAGAGGAA 6044

Qy       6157 ACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGTA 6216
Db       6045 GCCAAAAGAAACACAACCTTCTTTGCCCCTGACGGCGTGATGCATACGGTGCGCCATATA 6104

Qy       6217 TAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCAG 6276
Db       6105 CAACCGTTTGAGGTAGCCCTTGCGTGGAATATAGCGGAATGGCCAACGTTGATGCACCAG 6164

Qy       6277 GCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACTG 6336
Db       6165 CCCGTCGTGCACCATAAAATAGAGTAATCCATACGCCGTCATACCTGCGCCAATCCACTG 6224

Qy       6337 CAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACAC 6396
Db       6225 GAGCGGCCACATTCCTGTACTGCCCAGATAAATCAGCAGGATCGATAATGCAGCAAAAAC 6284

Qy       6397 CACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGATG 6456
Db       6285 CACGGCATAAAGATCGTTAACTTCAAACGCACCTTTACGCGGTTCATGATGTGAAAGATG 6344

Qy       6457 CCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCAT 6516
Db       6345 CCATCCCCAACCCCAGCCGTGCATGATGTATTTGTGTGCCAGTGCAGCAATCACTTCCAT 6404

Qy       6517 CACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATTGTTCGTCCATT 6576
Db       6405 GCCAATCACGGTAACGAAAACGATCAGGGCATTCCAAATCCACAACATAATTTCTCCGGT 6464

Qy       6577 TGTGGAAAAGGGAAGTA 6593
Db       6465 AGAGACGTCTGGCAGCA 6481
```

```
RESULT 2
ADQ14666
ID   ADQ14666 standard; DNA; 8609 BP.
XX
AC   ADQ14666;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Reporter plasmid pPCB15 nucleotide sequence SEQ ID NO:43.
XX
KW   carotenoid; carotenoid overproducing microorganism;
KW   functional isoprenoid enzymatic biosynthetic pathway; plasmid; circular;
KW   gene; ds.
XX
OS   Synthetic.
XX
PN   WO2004056974-A2.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-US041811.
XX
PR   19-DEC-2002; 2002US-0435612P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cheng Q,  Rouviere PE,  Tao L;
XX
DR   WPI; 2004-525436/50.
XX
PT   Novel carotenoid overproducing microorganism comprising genes encoding
PT   functional isoprenoid enzymatic biosynthetic pathway comprising disrupted
PT   genes such as deaD, mreC, and yfhE, useful for producing carotenoid.
XX
PS   Example 3; SEQ ID NO 43; 85pp; English.
XX
CC   The present invention describes a carotenoid overproducing microorganism
CC   comprising the genes encoding a functional isoprenoid enzymatic
CC   biosynthetic pathway comprising a disrupted gene chosen from deaD, mreC,
CC   and yfhE. Also described is a carotenoid overproducing Escherichia coli
CC   microorganism (II) comprising an upper isoprenoid enzymatic biosynthetic
CC   pathway comprising the genes dxs, dxr ygbP, ychB, ygbB, lytB, idi, ispA
CC   and ispB, a lower isoprenoid enzymatic biosynthetic pathway comprising
CC   the genes crtE, crtB, crtI, and crtY, mutations chosen from a mutation in
CC   the thrS gene of the 3159 nucleotide sequence of SEQ ID NO:35 (ADQ14658),
CC   mutation in the rpsA gene of the 2904 nucleotide sequence of SEQ ID NO:37
CC   (ADQ14660), mutation in the rpoC gene of the 5454 nucleotide sequence of
CC   SEQ ID NO:38 (ADQ14661), mutation in the yjeR gene of the 1845 nucleotide
CC   sequence of SEQ ID NO:39 (ADQ14662), or mutation in the rhoL gene of the
CC   2676 nucleotide sequence of SEQ ID NO:41 (ADQ14664), where the genes of
CC   the lower isoprenoid enzymatic biosynthetic pathway reside on an
CC   autonomously replicating plasmid comprising a replicon chosen from p15A
CC   and pMB1. (I) chosen from bacteria, yeasts and filamentous fungi such as
CC   Aspergillus, Trichoderma, Saccharomyces, Pichia, Candida, Hansenula,
CC   Salmonella, Bacillus, Acinetobacter, Zymomonas, Agrobacterium,
CC   Erythrobacter, Chlorobium, Chromatium, Flavobacterium, Cytophaga,
CC   Rhodobacter, Rhodococcus, Streptomyces, Brevibacterium, Corynebacteria,
CC   Mycobacterium, Deinococcus, Paracoccus, Escherichia, Erwinia, Pantoea,
CC   Pseudomonas, Sphingomonas, Methylomonas, Methylobacter, Methylococcus,
CC   Methylosinus, Methylomicrobium, Methylocystis, Alcaligenes,
CC   Synechocystis, Synechococcus, Anabaena, Thiobacillus, Staphylococcus,
CC   Methanobacterium, Klebsiella, and Myxococcus, and (II) are useful for
CC   producing carotenoids which involves contacting (I) or (II) with a
CC   fermentable carbon substrate, growing the carotenoid overproducing
CC   microorganism for a sufficient time to produce carotenoid, and optionally
CC   recovering the carotenoid from the carotenoid overproducing
CC   microorganism. The present sequence represents the reporter plasmid
CC   pPCB15 nucleotide sequence, which is used in an example from the present
CC   invention.
XX
SQ   Sequence 8609 BP; 1918 A; 2325 C; 2309 G; 2057 T; 0 U; 0 Other;

  Query Match             37.4%;  Score 2619.6;  DB 12;  Length 8609;
```

```
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy       373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
             ||| ||| |     | |  |    |  ||||| || |||||||| |         |  |
Db      1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy       433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
             ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||
Db      1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy       493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
              |  | ||| |  |  ||||||||||| | || || | | | |  ||    ||  | ||
Db      1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy       553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
             ||| ||||| || || ||||||||||| || || ||||||||||||||||| || |||
Db      1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy       613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
             |  |||||  | |||||||||||||| || || || ||| | ||     | |  ||||||
Db      1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy       673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
              | ||||||||||| || ||  ||||  | || |||   || ||||  |||||||||
Db      1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy       733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
             ||    ||| || |  | |||   ||| ||   ||| |  | || || ||| ||||  |
Db      1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy       793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
             ||| |||||||  | || || || || ||  |||| ||    || |||    |   ||||
Db      2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy       853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
             ||| | || || ||     | |||||  |  | |||||||||  ||||||| ||||  |
Db      2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy       913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
               |||||||||| | ||||| ||  |||| ||     | | |||  | |  ||
Db      2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy       973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
             ||| | |  |||| | |||||||| |||||  |||| ||||| ||  |||| ||    |
Db      2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy      1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
              | ||||| |||||  |  | ||||||   || ||||| ||||||||  |  |  | ||
Db      2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy      1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
              | | ||  |  |||   ||||||||  | ||  |||||   ||   || ||||||| ||
Db      2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406

Qy      1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
                || |||||   ||||    ||||   ||||   | ||| || ||   |||| |  ||
Db      2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy      1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
              | || ||     |||  | ||   |   ||     || |||| |||||||||||||| |
Db      2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy      1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
             ||||  | | ||||||  ||| |||  ||||     | || ||     |  |||| ||||
Db      2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy      1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
             | |||||  | || ||  | ||||||   ||   |    |  || | |  | |      |
Db      2582 GTCATCGTGTTACGTTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641

Qy      1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
             |||  ||     ||||||| |  || ||||||||| ||||| |  |  ||| | || |||
Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061

Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
Db       3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
Db       3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
Db       3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTTCCGTCCCCGGACA 3241

Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
Db       3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
Db       3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
Db       3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
Db       3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGGATGAATA 3481

Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
Db       3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
Db       3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy       2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
Db       3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy       2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
Db       3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy       2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
Db       3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780
```

```
Qy       2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
Db       3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy       2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
Db       3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy       2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
Db       3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy       2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
Db       3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy       2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
Db       4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy       2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
Db       4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy       2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
Db       4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy       2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
Db       4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy       3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
Db       4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy       3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
Db       4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy       3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
Db       4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy       3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
Db       4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy       3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
Db       4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy       3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
Db       4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy       3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
Db       4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy       3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
Db       4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy       3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
Db       4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy       3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
Db       4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857
```

```
Qy    3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
           |||| ||||||||||||| || ||| |    |  ||  |||  ||  |  |||   ||  
Db    4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy    3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
                 | |   ||  | |||||||  |||   || ||||| || |||||||| ||||||
Db    4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy    3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
           || ||||| |||||  | || || || || || ||        | || ||||||||||||
Db    4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy    3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
           || ||||| || || || |||| | | ||||    |||||||| || ||||| || || 
Db    5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy    3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
           || || |||||||||||||||||| || |||||  ||||  | ||||| || ||   ||| 
Db    5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy    3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
           |  ||||||||||||||| |||||||| ||||| || |||||||||||||||||    |||
Db    5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy    3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
           ||     |  ||||||| ||  | | | | | | | ||   ||||||      ||| |||
Db    5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy    4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
           ||||  ||||| |  || |||||    || |||| ||||||| ||||  ||||| |  |||
Db    5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy    4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
           || ||||||||| |||||||| ||||| ||| |   | |  |  |||||||| || | ||||
Db    5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy    4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
           || ||||||| |   ||||| ||||||||||||||| ||||| |  || ||    |  |||
Db    5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy    4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
            |||| || ||||| || ||||||||||||| |||||  |  | |||||  || ||||| 
Db    5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGGAATCCG 5517

Qy    4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
           |||||||| || || || |||  | |||||| ||| | || || ||||||||||| |||
Db    5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy    4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
           ||||| ||||| || ||||| ||||||||  | |||||       |||||  ||||  ||
Db    5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy    4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
           ||||||||  |  | || || |||   || || ||  ||||||||   ||    |   ||
Db    5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy    4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
              | ||| ||||| || | |||  |||| || ||  || || ||| ||||| || || || 
Db    5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757

Qy    4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
           || || || |||||||||   | | |||||    || |||||    ||    || |  |||
Db    5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy    4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
           | |   ||| |    ||||| ||||| ||||| ||||||||||||||||||||| || || 
Db    5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy    4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
           || |  || || || ||||| || || ||||||||||||| || || ||  |||||  ||
Db    5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy    4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
```

```
           ||  | || ||||| |||         ||||| || |||||||| |||||  | ||||| 
Db    5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy    4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
           || ||   ||  |||||||| |||||||||   || |||||||||| ||||||| | |||
Db    5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy    4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
           || || || ||    ||||| ||  || ||||||||      ||||||||  ||  |||||
Db    6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy    4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
           ||||| |||||||  || || ||||| || |||||||| || || ||    || || |||
Db    6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy    4927 ACACACAGAATGTTTACGCCGTTTGATTTCCGCGACACGCTGCATGCCCATCACGGCTCG 4986
           || ||| | ||||||||||||| ||||||||||||  |||  |||||   || || |||
Db    6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy    4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
           || || ||| | || || ||| |||| || ||||||||||||||| || ||||||||||||
Db    6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy    5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
               | ||    ||||| ||||||| || || || || || ||||| ||||||||| | |||
Db    6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy    5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
           ||||| ||||| ||||| ||||| || || ||  |||| |||||||||||         | ||
Db    6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy    5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
           ||     | | |||||  | || |      ||||||||||||| |||||||| || || |
Db    6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy    5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
           | || ||  | | ||| || || ||    || || |||||||  |||||||||| || || |
Db    6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy    5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
           |||| || |||||||| || || ||||| |  ||||| ||||||      | |   ||||
Db    6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy    5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
              |     |     |  ||   ||  ||||||| ||  ||||  |||| || |  ||||
Db    6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy    5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
           ||  |||  |||| ||| | || || || |||   |||||||||||| || |||||    |
Db    6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy    5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
           |  |  |  | |        ||||| || |||||||||||| || || |||||||  ||| 
Db    6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774

Qy    5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
           |    |  |||  ||   | || || ||||||||| ||||||||||||||| || || |
Db    6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy    5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
           | ||  ||||||||||||   | | ||||||| | |||||  | ||   |||||||| |||||
Db    6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy    5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
           ||||| | || ||||| ||||||| | || | ||||||||||| || ||||| || |||||  
Db    6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy    5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
           |    || |||| ||||||||  ||  ||||| ||  |||  ||          |  |
Db    6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014

Qy    5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                 |||| ||  || | ||    ||  |   |    | || | |||  | ||   ||
Db        7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy        5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
               | ||||||||||||||    ||  |     ||||| |  |    |||||  | |||||||
Db        7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy        5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
               | ||||||||||| || ||       || || ||  |||||||||| |||||| | || |
Db        7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy        5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
               |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db        7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy        6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
               || |||  |   | |||  |||  | ||||| ||  ||        |  ||| ||||| |
Db        7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy        6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
               ||  ||| |||||||| ||        |  ||    | || |  |  | |||||| || ||
Db        7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy        6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
                || |||   | |||||||||||| || |||||||| ||||||||||| ||    || |
Db        7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy        6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
                || ||||||||||| ||||| ||||||||||| || || || |||||||||||||| ||
Db        7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy        6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
                ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db        7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy        6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
                || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db        7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy        6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
               ||||||  |||||| ||  |  || || ||||| ||||| |||||||| |  |||||
Db        7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy        6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
                ||  |  |||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db        7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy        6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
               |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db        7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy        6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
                ||||| | ||| | |||  ||  || |
Db        7789 ATTCCAAATCCACAACATAATTTCTCAA 7816
```

```
RESULT 3
ADQ26394
ID   ADQ26394 standard; DNA; 8609 BP.
XX
AC   ADQ26394;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Carotenoid reporter plasmid pPCB15.
XX
KW   Chromosomal engineering; carotenoid; pPCB15; ds.
XX
OS   Unidentified.
XX
PN   WO2004056973-A2.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-US041810.
XX
PR   19-DEC-2002; 2002US-0434602P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Suh W;
XX
DR   WPI; 2004-507710/48.
XX
PT   Directed integration of an expressible DNA fragment lacking a selectable
PT   marker into a bacterial chromosome comprises co-transforming
PT   recombination proficient host with at least two linear recombination
PT   elements.
XX
PS   Disclosure; SEQ ID NO 37; 96pp; English.
XX
CC   The invention provides a method for the integration of foreign genetic
CC   elements into a bacterial chromosome without the need of a cloning step.
CC   The method relies on the presence of the bacteriophage lambda-Red
CC   recombinase system in a recombination proficient bacterial host. At least
CC   2 linear recombination elements or constructs are co-transformed into the
CC   host, and are assembled in the correct orientation in the bacterial
CC   chromosome by homologous recombination. Selectable markers used for the
CC   selection of transformants are later excised by the action of a site-
CC   specific recombinase. The method allows for multiple chromosomal
CC   modifications within Escherichia coli, which is essential when
CC   engineering biosynthetic pathways for industrial purposes. The method is
CC   illustrated using isoprenoid/carotenoid biosynthesis as an example. The
CC   promoters of key genes that encode the rate-limiting enzymes involved in
CC   the isoprenoid pathway are engineered, resulting in increased beta-
CC   carotene production. The present sequence is that of carotenoid reporter
CC   plasmid pPCB15, which was used in examples from the invention.
XX
SQ   Sequence 8609 BP; 1918 A; 2325 C; 2309 G; 2057 T; 0 U; 0 Other;

  Query Match             37.4%;  Score 2619.6;  DB 12;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy         373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
               ||| ||| |     | |  |    |  ||||| || |||||||| |         |  |
Db        1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy         433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
               ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||
Db        1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy         493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
                |  | ||| |  |  ||||||||||| | || || | | | |  ||    ||  | ||
Db        1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy         553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
               ||| | ||||| || || ||||||||||| || || ||||||||||||||||| || |||
Db        1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy         613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
               |  |||||  | |||||||||||||| || || || ||| | ||     | |  ||||||
Db        1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy         673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
                | ||||||||||| || ||  ||||  | || |||   || ||||  |||||||||
Db        1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy         733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
               ||    |||  || |  | |||   ||| ||   ||| |  | || || ||| ||||  |
Db        1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy         793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
               ||| |||||||  | || || || || ||  |||| ||    || |||    |   ||||
Db        2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy         853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
```

```
Db       2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972

Db       2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032

Db       2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092

Db       2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152

Db       2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406

Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209

Db       2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269

Db       2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329

Db       2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389

Db       2582 GTCATCGTGTTACGTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641

Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449

Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509

Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569

Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629

Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689

Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749

Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809

Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061

Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869

Db       3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929

Db       3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
```

```
Db       3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTTCCGTCCCCGGACA 3241

Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046

Db       3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGAT 2106

Db       3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166

Db       3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226

Db       3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGGATGAATA 3481

Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286

Db       3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346

Db       3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy       2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406

Db       3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy       2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466

Db       3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy       2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526

Db       3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy       2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586

Db       3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy       2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646

Db       3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy       2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706

Db       3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy       2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766

Db       3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy       2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826

Db       4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy       2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886

Db       4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy       2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946

Db       4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy       2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006

Db       4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy       3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066

Db       4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320
```

```
Qy      3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
Db      4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy      3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
Db      4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy      3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
Db      4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy      3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
Db      4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy      3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
Db      4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy      3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
Db      4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy      3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
Db      4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy      3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
Db      4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy      3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
Db      4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy      3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
Db      4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy      3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
Db      4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy      3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
Db      4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy      3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
Db      5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy      3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
Db      5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy      3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
Db      5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy      3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
Db      5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy      4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
Db      5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy      4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
Db      5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy      4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
Db      5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy      4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
Db      5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGGAATCCG 5517

Qy      4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
Db      5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy      4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
Db      5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy      4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
Db      5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy      4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
Db      5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757

Qy      4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
Db      5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy      4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
Db      5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy      4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
Db      5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy      4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
Db      5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy      4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
Db      5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy      4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
Db      6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy      4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
Db      6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy      4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
Db      6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy      4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
Db      6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy      5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
Db      6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy      5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
Db      6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy      5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
Db      6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy      5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
```

```
              | || ||  | | ||| || || ||    || || ||||||  ||||||||| || || |
Db       6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy       5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
              |||| || |||||||| || || ||||| |  ||||| ||||||      | |   ||||
Db       6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy       5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
                 |     |     |  ||   ||  ||||||| ||  ||||  |||| || |  ||||
Db       6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy       5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
              ||  |||  |||| ||| | || || || |||   |||||||||||| || |||||    |
Db       6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy       5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
              |  |  |  | |        ||||| || ||||||||||| || || |||||||  |||
Db       6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774

Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
              |    |  |||  ||   | || || ||||||||||| ||||||||||||| || || |
Db       6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
              | ||  ||||||||||||   | ||||||||| |||||  | ||   ||||||||| |||||
Db       6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
              ||||| | || ||||| |||||| | || |||||||||||| || ||||| || |||||
Db       6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
              |    || ||||| ||||||||  ||  ||||||| ||  |||  ||| |       |  |
Db       6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                    |||| ||  || | ||    ||  |   |    | || | |||  | ||   ||
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
              | ||||||||||||||    ||  |     ||||| |  |    |||||  | |||||||
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | ||||||||||| || ||       || || ||  |||||||||| |||||| | || |
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              || |||  |   | |||  |||  | ||||| ||  ||        |  ||| ||||| |
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
              ||  ||| |||||||| ||        | ||    | || |  |  | ||||||| || ||
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
               || |||   | ||||||||||||| || ||||||| ||||||||||| ||    || |
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy       6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
               || ||||||||||| ||||| |||||||||||| || || || |||||||||||||| ||
Db       7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy       6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
               ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db       7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy       6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
               || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db       7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy       6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
              ||||||  |||||| ||  |  || || ||||| ||||| |||||||| |  |||||
Db       7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy       6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
               ||  |  ||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db       7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy       6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
              |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db       7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy       6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
               ||||| | ||| | |||  ||  || |
Db       7789 ATTCCAAATCCACAACATAATTTCTCAA 7816


RESULT 4
ADQ77262
ID   ADQ77262 standard; DNA; 8609 BP.
XX
AC   ADQ77262;
XX
DT   23-SEP-2004  (first entry)
XX
DE   Nucleotide sequence of plasmid pPCB15.
XX
KW   carotenoid; carotenoid enzymatic biosynthetic pathway; dxs gene;
KW   D-1-deoxyxylulose 5-phosphate gene; idi gene; ygbBP gene; ispB gene;
KW   yjeR gene; oligo-ribonuclease gene; dxs gene; dxr gene;
KW   DOXP reductoisomerase; ygpP gene; ispD gene; ychB gene; ispE gene;
KW   ygbB gene; ispF gene; lytB gene; ispH gene; idi gene;
KW   isopentenyl diphosphate isomerase; ispA gene;
KW   farnesyl pyrophosphate synthase; ispB gene;
KW   octaprenyl diphosphate synthase; crtE gene;
KW   geranylgeranyl pyrophosphate synthase; crtB gene; phytoene synthase;
KW   crtl gene; phytoene dehydrogenase; crtX gene;
KW   zeaxanthin glucosyl transferase; crtZ gene; beta-carotene hydroxylase;
KW   crtY gene; lycopene cyclase; enzyme; ss.
XX
OS   Synthetic.
XX
PN   WO2004056975-A2.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-US041812.
XX
PR   19-DEC-2002; 2002US-0434618P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cheng Q,  Rouviere PE,  Suh W;
XX
DR   WPI; 2004-525437/50.
XX
PT   Novel carotenoid overproducing bacteria such as Escherichia coli or
PT   Flavobacterium, comprising genes encoding functional carotenoid enzymatic
PT   biosynthetic pathway, useful for producing carotenoids such as
PT   astaxanthin and beta-carotene.
XX
PS   Disclosure; SEQ ID NO 64; 137pp; English.
XX
CC   The specification describes a carotenoid overproducing bacteria, which
CC   comprise genes encoding a functional carotenoid enzymatic biosynthetic
CC   pathway, where the dxs (D-1-deoxyxylulose 5-phosphate gene), idi, ygbBP
CC   and ispB genes are overexpressed, and the yjeR gene (oligo-ribonuclease
CC   gene) is down regulated. In bacteria of the invention, the carotenoid
CC   enzymatic biosynthetic pathway consists of the genes dxs, dxr (DOXP
CC   reductoisomerase), ygpP (also known as ispD), ychB (also known as ispE),
CC   ygbB (also known as ispF), lytB (also known as ispH), idi (isopentenyl
CC   diphosphate isomerase), ispA (farnesyl pyrophosphate synthase), ispB
CC   (octaprenyl diphosphate synthase), crtE (geranylgeranyl pyrophosphate
CC   synthase), crtB (phytoene synthase), crtl (phytoene dehydrogenase), crtX
CC   (zeaxanthin glucosyl transferase), crtZ (beta-carotene hydroxylase) and
CC   crtY (lycopene cyclase). The lytB and dxr gene are optionally
CC   overexpressed. Carotenoid overproducing bacteria of the invention are
CC   useful for producing carotenoids. The present sequence represents a
CC   plasmid pPCB15, which is used during construction of bacteria of the
CC   invention.
XX
SQ   Sequence 8609 BP; 1918 A; 2325 C; 2309 G; 2057 T; 0 U; 0 Other;

  Query Match             37.4%;  Score 2619.6;  DB 12;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy        373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
Db       1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686
Qy        433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
Db       1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746
Qy        493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
Db       1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806
Qy        553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
Db       1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866
Qy        613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
Db       1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926
Qy        673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
Db       1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986
Qy        733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
Db       1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046
Qy        793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
Db       2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106
Qy        853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
Db       2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166
Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
Db       2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226
Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
Db       2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286
Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
Db       2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346
Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
Db       2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406
Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
Db       2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466
Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269

Db       2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521
Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
Db       2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581
Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
Db       2582 GTCATCGTGTTACGTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641
Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701
Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761
Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821
Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881
Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941
Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001
Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061
Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
Db       3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121
Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
Db       3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181
Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
Db       3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTCCGTCCCCGGACA 3241
Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
Db       3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301
Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGAT 2106
Db       3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361
Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
Db       3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421
Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
Db       3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGATGAATA 3481
Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
Db       3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541
Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
```

```
Db        3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy        2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
                       |||   |||| || |||  | | |  |||||||||  ||     |||| |||
Db        3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy        2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
                 || ||| ||   ||||||||     |||  |  ||||||||    |     |||||||
Db        3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy        2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
               | || || || |||||| || |  | | |  || |        |  |||    | ||||
Db        3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy        2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
                 |||| || ||| | ||| |||||||||| || ||||| |||||| | || ||||| ||
Db        3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy        2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
                || |  |||  ||| ||  |  ||     | || |  | |||   | | |  |  |||
Db        3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy        2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
               ||| || ||||||||||| |||||  |    ||| | |      | || | ||||||| |
Db        3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy        2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
                   |||||  |   |   |  ||| | |  || ||||| ||||| ||    |  || ||
Db        3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy        2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
               | || || || | |  || |  | | |  | ||  |  |||| ||| |    ||
Db        4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy        2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
                  | | ||  |  || | |||    | || || ||   |   ||  |  |      |||
Db        4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy        2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
                  | | |||||||||||     ||  |||       ||||| || || || ||
Db        4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy        2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
               |||     |    || |   | || |  ||||  ||| | || || |||||||| |||
Db        4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy        3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
                  |||||| ||  |   |   |||||  | |||||||| || ||||||||||||   ||
Db        4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy        3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
                || || ||||| || || ||||||||   |||       ||  |||| ||||| || ||
Db        4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy        3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
                ||| || || || |    ||   ||   |    || |||||| ||||   ||| |||||
Db        4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy        3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
                   |  || || ||| || |    ||| ||||||| ||||| || ||   ||| || ||
Db        4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy        3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
               |||  | |  ||| | |    ||  || ||||||||||| |     ||||| || |||||
Db        4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy        3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
                 | |||||||||||||||||| | || ||||| |||||| | |||  ||| ||||||||
Db        4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy        3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
                || ||     |    ||||||             |   | |  |   |||    || |
Db        4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677
```

```
Qy        3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
                | ||||||||  || |  ||||   ||||| | ||||||||  |||| || |||||| |
Db        4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy        3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
               |||||| |||||   | |  ||   |||||||| |||||||| ||||| ||  |  |
Db        4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy        3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
                |  |  || |||  ||||||||| || ||  ||||   | ||  ||||      || ||
Db        4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy        3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
               |||| ||||||||||||| || ||| |    |  ||  |||  ||  |  |||   ||
Db        4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy        3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
                     | |   ||  | |||||||  |||   || ||||| || |||||||| ||||||
Db        4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy        3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
               || ||||| |||||  | || || || || || ||        | || ||||||||||||
Db        4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy        3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
               || ||||| || || || ||||| | | ||||    |||||||| || ||||| || ||
Db        5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy        3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
               || || ||||||||||||||||| || |||||  ||||  | ||||| || |||  |||
Db        5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy        3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
               |  ||||||||||||||| ||||||| ||||| || |||||||||||||||||    |||
Db        5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy        3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
               ||     |  ||||||| ||  | ||| | | | | ||   ||||||      ||| |||
Db        5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy        4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
               ||||  ||||| |  || |||||    || |||| |||||| ||||  ||||| |  |||
Db        5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy        4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
               || |||||||| |||||||| ||||| ||| |   | |    |||||||| || | ||||
Db        5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy        4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
               || ||||||| |   ||||| |||||||||||||| ||||| |  || ||    |  |||
Db        5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy        4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
                |||| || ||||| || ||||||||||| || |||||  |  | ||||| || |||||
Db        5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGGAATCCG 5517

Qy        4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
               |||||||| || || || |||||  | ||||| ||| | || || ||||||||||| |||
Db        5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy        4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
               ||||| ||||| || ||||| ||||||||  | ||||||      |||||  ||||  ||
Db        5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy        4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
               ||||||||  |  | || || |||   || || ||  ||||||||   ||  |    |||
Db        5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy        4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
                  | ||| ||||| || | |||  |||| || ||  ||  ||| ||||| || || ||
Db        5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757
```

```
Qy       4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
Db       5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy       4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
Db       5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy       4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
Db       5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy       4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
Db       5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy       4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
Db       5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy       4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
Db       6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy       4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
Db       6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy       4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
Db       6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy       4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
Db       6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy       5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
Db       6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy       5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
Db       6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy       5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
Db       6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy       5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
Db       6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy       5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
Db       6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy       5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
Db       6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy       5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
Db       6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy       5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
Db       6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774

Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
Db       6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
Db       6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
Db       6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
Db       6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy       6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
Db       7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy       6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
Db       7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy       6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
Db       7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy       6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
Db       7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy       6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
Db       7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy       6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
Db       7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy       6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
Db       7789 ATTCCAAATCCACAACATAATTTCTCAA 7816
```

```
RESULT 5
ADQ48673
ID   ADQ48673 standard; DNA; 8609 BP.
XX
AC   ADQ48673;
```

```
XX
DT   23-SEP-2004  (first entry)
XX
DE   Plasmid pPCB15 DNA containing Pantoea stewartii crtEXYIB genes.
XX
KW   homologous recombination system; biosynthetic pathway design;
KW   metabolic flux optimisation; carotenoid cluster; plasmid pPCB15; ds;
KW   crtEXYIB.
XX
OS   Pantoea stewartii.
OS   Unidentified.
OS   Synthetic.
XX
PN   WO2004056972-A2.
XX
PD   08-JUL-2004.
XX
PF   19-DEC-2003; 2003WO-US041678.
XX
PR   19-DEC-2002; 2002US-0434773P.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Rouviere PE,  Suh W;
XX
DR   WPI; 2004-507709/48.
XX
PT   Optimizing production of a genetic end product comprises utilizing a
PT   phage transducing system to transfer multiple integration cassettes into
PT   a single host cell in parallel combinatorial fashion.
XX
PS   Example 6; SEQ ID NO 40; 103pp; English.
XX
CC   The invention relates to a novel method of optimising the production of a
CC   genetic end product which comprises utilising a homologous recombination
CC   system to introduce an integration cassette into a chromosome of a
CC   recombination proficient host cell and subsequently utilising a phage
CC   transducing system to transfer the multiple integration cassettes into a
CC   single host cell in parallel combinatorial fashion. The method of the
CC   invention may be useful for redesigning biosynthetic pathways, optimising
CC   metabolic flux and creating novel pathways by targeted in vivo
CC   chromosomal engineering. The current sequence is that of the plasmid
CC   pPCB15 DNA of the invention which contains the Pantoea stewartii
CC   carotenoid cluster crtEXYIB genes.
XX
SQ   Sequence 8609 BP; 1918 A; 2325 C; 2309 G; 2057 T; 0 U; 0 Other;

  Query Match             37.4%;  Score 2619.6;  DB 12;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy        373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
              ||| ||| |     | |  |    |  ||||| || |||||||| |         |  |
Db       1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy        433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
              ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||
Db       1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy        493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
               |  | ||| |  |  |||||||||| | || || | | | |  ||    ||  | ||
Db       1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy        553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
              ||| | ||||| || || ||||||||| || || |||||||||||||||||| || |||
Db       1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy        613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
              |  |||||  | |||||||||||||| || || || ||| | ||     | |  ||||||
Db       1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy        673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
               | ||||||||||| || ||  ||||  | || |||   || ||||  |||||||||
Db       1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy        733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
              ||    ||| || |  | ||| |   ||| ||   ||| |  | || || ||| ||||  |
Db       1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy        793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
              ||| |||||||  | || || || || ||  |||| ||     || |||     |   ||||
Db       2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy        853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
              ||| | || || ||     | |||||  |  | |||||||||  |||||||| ||||  |
Db       2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
                |||||||||| | ||||| ||  |||| ||     | | |||  | | |  ||
Db       2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
              ||| | |  ||| | |||||||| ||||| |||| |||| ||  |||| ||    |
Db       2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
               | ||||| |||||  |  | ||||||   || ||||| ||||||||  |  |  |  ||
Db       2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
               | | ||  |  |||   ||||||||  | ||  |||||   ||   || ||||||| ||
Db       2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406

Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
                 || |||||   ||||    ||||   ||||||   | ||| || ||   |||| |  ||
Db       2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
               | || ||     |||  | ||   |   ||     || |||| |||||||||||||| |
Db       2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
              ||||  | | ||||||  ||| |||  ||||     | || ||     |  |||| ||||
Db       2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
              | |||||  | || ||  | ||||||   ||   |    |  || | |  | |      |
Db       2582 GTCATCGTGTTACGTTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641

Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
              |||  ||     ||||||| |  || ||||||||| |||||||  |  ||| | || |||
Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
              | | |   ||    ||   ||| |  ||| |||| ||  ||||| | ||   | ||
Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
               || |||||||||| ||||| |||||||| || |   |  |  || ||| | || || ||||
Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
              ||||||  ||  |||||||   |||| || | | | | |||||  ||||||     ||||||||
Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
              |||||||||||||||||||  ||||| ||||| || ||| | ||  | || || |||||||
Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
              |||| |||   |  |        ||    || | | ||||  | |    |||||    | |
Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
              | |||||  |  | ||||| |||||||     ||||||    | || ||  |     |
Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061
```

```
Qy   1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
           |||  |   |||      | | ||||| ||  | || || |||||||||||||||||||| 
Db   3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy   1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
          || | || |   | |||||||| ||  |    ||||| | ||||| ||| || || || |
Db   3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy   1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
          |  | ||     ||| ||   | || | | |      | |    |  ||   | | |    
Db   3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTTCCGTCCCCGGACA 3241

Qy   1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
          | ||     ||| |||||||||||||| || ||||| |||| || ||| |   | |||||   | 
Db   3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy   2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
             | ||  | || |||    || ||        | |  || | || || ||||| ||  
Db   3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy   2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
          |    |||   ||| |   | | ||||  |  |  |  || |       |||   |||||| |
Db   3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy   2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
           ||||||   |||||| ||    || || ||| || ||   |||||| ||||  ||| | || |
Db   3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGGATGAATA 3481

Qy   2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
            |  ||||| ||  | |  | | | || ||| | ||||||||||||| | || |||||||
Db   3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy   2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
          | || ||||||||  | || ||||  |  ||||| | | ||    || ||  || || ||||
Db   3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy   2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
                  |||   |||| || ||  | | |  ||||||||||  ||     |||| |||
Db   3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy   2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
            || ||| ||   ||||||||     |||  |  |||||||||    |     |||||||
Db   3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy   2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
          | || || || |||||| || |  | | |  || |        |  |||    | |||| 
Db   3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy   2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
            |||| || ||| | ||| |||||||||| || ||||| |||||| | | || ||||| ||
Db   3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy   2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
           || |  |||  ||| ||  |  ||     | || |  | |||   | | |  |  ||| 
Db   3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy   2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
          ||| || |||||||||||| |||||  |    ||| | |      | || | ||||||| |
Db   3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy   2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
              |||||  |   |   |  ||| | |  || ||||| ||||| ||    |  || ||
Db   3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy   2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
          | || || || | |  || |  | | |  | ||  |  ||  |||| || ||| |    ||      
Db   4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy   2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
             | | ||  |  || | |||    | || || ||   |   ||  |  |      |||
Db   4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy   2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
             | | ||||||||||     ||  |||       ||||| || || || |  | ||      
Db   4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy   2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
          |||     |    || |   | || |  ||||  ||| | || || |||||||| |||  
Db   4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy   3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
             |||||| ||  |   |   |||||  | |||||||| || |||||||||||||   ||
Db   4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy   3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
           || || ||||| || || |||||||||   |||       ||  |||| |||| || ||
Db   4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy   3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
           ||| || || || |    ||   ||   |    || ||||||| ||||   ||| |||||
Db   4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy   3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
              |  || || ||| || |    ||| |||||| ||||| || ||   ||| || ||
Db   4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy   3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
          |||  | |  ||| | |    ||  || ||||||||||| |     ||||| || |||||
Db   4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy   3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
            | |||||||||||||||||| | || ||||| |||||| | |||  ||| ||||||||
Db   4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy   3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
           || ||     |    ||||||                |   | |  |   |||    || | 
Db   4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy   3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
           | ||||||||  || |  ||||   |||||| | ||||||||  |||| || |||||| |
Db   4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy   3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
          |||||| |||||   | |  ||   |||||||| ||||||||| ||||| ||  |  |   
Db   4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy   3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
           |  |  || |||  |||||||||| || ||  ||||   | ||  ||||      || ||
Db   4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy   3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
          |||| ||||||||||||| || ||| |    |  ||  ||  |  ||  |  |||   ||  
Db   4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy   3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
                | |   ||  | |||||||  |||   || |||| | || ||||||||| ||||||
Db   4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy   3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
          || ||||| |||||  | || || || || || ||        | || ||||||||||||
Db   4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy   3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
          || ||||| || || || ||||| | | ||||    |||||||| || ||||| || || 
Db   5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy   3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
          || || |||||||||||||||||| || |||||  ||||  | ||||| ||  |||  ||| 
Db   5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy   3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGAAGATGGC 3966
          |  ||||||||||||||| |||||| |||||| || ||||||||||||||||||    |||
Db   5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy   3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
```

```
Db       5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy       4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
Db       5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy       4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
Db       5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy       4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
Db       5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy       4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
Db       5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGAATCCG 5517

Qy       4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
Db       5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy       4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
Db       5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy       4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
Db       5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy       4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
Db       5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757

Qy       4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
Db       5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy       4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
Db       5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy       4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
Db       5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy       4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
Db       5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy       4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
Db       5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy       4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
Db       6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy       4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
Db       6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy       4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
Db       6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy       4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
Db       6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy       5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
```

```
Db       6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy       5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
Db       6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy       5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
Db       6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy       5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
Db       6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy       5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
Db       6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy       5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
Db       6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy       5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
Db       6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy       5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
Db       6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774

Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
Db       6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
Db       6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
Db       6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
Db       6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428
```

```
Qy      6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
              || ||||||||||| ||||| ||||||||||| || || || |||||||||||||| ||
Db      7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy      6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
              ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db      7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy      6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
              || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db      7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy      6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
             ||||||  |||||| ||  |  || || ||||| ||||| |||||||| |  |||||
Db      7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy      6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
              ||  |  |||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db      7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy      6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
             |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db      7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy      6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
              ||||| | ||| | |||  ||  || |
Db      7789 ATTCCAAATCCACAACATAATTTCTCAA 7816


RESULT 6
ADM98599
ID   ADM98599 standard; DNA; 12753 BP.
XX
AC   ADM98599;
XX
DT   01-JUL-2004  (first entry)
XX
DE   Geranylgeranyl pyrophosphate synthase DNA #19.
XX
KW   Geranylgeranyl pyrophosphate synthase; gene; ds; diterpene;
KW   diterpene precursor; diterpene synthase; defence toxin;
KW   volatile defensive signal; pollinator attractant; photoprotectant.
XX
OS   Pantoea agglomerans.
XX
PN   US2004072323-A1.
XX
PD   15-APR-2004.
XX
PF   07-JAN-2002; 2002US-00041018.
XX
PR   05-JAN-2001; 2001US-0259880P.
XX
PA   (MATS/) MATSUDA S P T.
PA   (HART/) HART E A.
XX
PI   Matsuda SPT,  Hart EA;
XX
DR   WPI; 2004-373921/35.
XX
PT   New unicellular organisms comprising exogenous nucleic acids encoding a
PT   geranylgeranyl pyrophosphate and a diterpene synthase, useful for
PT   producing diterpenes and diterpene precursors.
XX
PS   Claim 2; SEQ ID NO 19; 38pp; English.
XX
CC   The invention relates to a unicellular organism for producing a diterpene
CC   or diterpene precursor comprising an exogenous nucleic acid sequence
CC   encoding a geranylgeranyl pyrophosphate synthase under the control of a
CC   promoter operable in the organism, and an exogenous nucleic acid sequence
CC   encoding a diterpene synthase under the control of a promoter operable in
CC   the organism. The invention also relates to methods of producing a
CC   diterpene or diterpene precursor and a method of isolating a diterpene
CC   synthase comprising growing several cells in the presence of a
CC   polyaromatic resin to make a cell/resin mixture, where at least one of
CC   the cells further comprises at least one isolated and purified nucleic
CC   acid sequence of a yeast expression library, and the expression of the
CC   nucleic acid sequence is regulated by an inducible promoter under
CC   conditions where the expression is induced, filtering the cell/resin
CC   mixture, extracting the cell/resin mixture with alcohol to produce an
CC   organic eluent and analysing the organic eluent by a screening method
CC   including chromatography and/or spectroscopy, to identify the nucleic
CC   acid sequence encoding the diterpene synthase. The unicellular
CC   microorganism is useful as a diterpene or diterpene precursor producing
CC   system. Diterpenes, in plants, serve as defence toxins, volatile
CC   defensive signals, pollinator attractants and photoprotectants. This
CC   sequence represents DNA encoding a geranylgeranyl pyrophosphate synthase
CC   polypeptide of the invention. Note: The sequence data for this patent did
CC   not form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 12753 BP; 2489 A; 3880 C; 3911 G; 2473 T; 0 U; 0 Other;

  Query Match             30.5%;  Score 2135.2;  DB 12;  Length 12753;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 3263;  Conservative    0;  Mismatches 1765;  Indels   39;  Gaps    4;

Qy      1479 CTGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCT 1538
             ||| | ||  |||| |  |   |  | |||   |||||| |||| ||||   |  ||||
Db      5780 CTGCTGCGGCTGATTGCGGCCATGTCATCCCTGACCGATGTGCTCTGCCAGCAGTTGCCC 5839

Qy      1539 GCGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGC 1598
             ||  | ||  |     ||    | || |   |||| | ||| || |||||  | || ||
Db      5840 GCTATTCTACAGCGGCTGGCGGTGGACGCGCTGATTGTCGATGAGATGGAGCCCGCCGGA 5899

Qy      1599 GGATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTC 1658
              |  ||||||| || ||||||   || || ||| | ||  | ||||||||  ||||||||
Db      5900 AGCCTGGTCGCCGAGGCGCTGGGACTACCATTTATCTCTATTGCCTGCGCGCTGCCGGTC 5959

Qy      1659 AATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAA 1718
             || || ||  | ||  | |||||  |||||||||| || |  |  ||  ||||| | | |
Db      5960 AACCGCGAGCCGGGTCTGCCGCTGCCGGTGATGCCGTTTCACTACGCCGAGGATAAGAGA 6019

Qy      1719 GCGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGC 1778
             || |||   ||||||||||| ||||      ||||| |||    | |||    |  |||||
Db      6020 GCCCTGCGGCGTTTTCAGGTCAGCGAACGGATCTACGATGCGCTGATGTACCCGCACGGG 6079

Qy      1779 GACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAG 1838
              |   ||||||    |||||| | |   |||  |||| |||||||  |  |  |  |  ||
Db      6080 CAGACGATCCTGCGCCACGCCCAGCGCTTTGGTTTGCCGGAGCGCAGGCGTCTCGACGAG 6139

Qy      1839 TGCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAG 1898
             || || ||||||||||| || || ||||||   || |||||| | || || ||||| | |
Db      6140 TGTCTCTCGCCGCTGGCGCAGATTAGCCAGTCCGTTCCGGCCCTCGACTTCCCACGCCGG 6199

Qy      1899 CAACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCG 1958
                |||||   ||| | ||||  ||||||  |||| |||         | ||   |||||
Db      6200 GCGCTGCCGAACTGTTTTCACTACGTGGGAGCACTGCGC---------TATCAGCCGCCG 6250

Qy      1959 CTCCATGCGCCCTGGCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTG 2018
             |  || |      |  |  | |   |   ||||  | | | |||||||||||| || ||
Db      6251 CCGCAGGTAGAACGCTCGCCACGCAGCACGCCGCGGATCTTTGCCTCGCTGGGCACCCTC 6310

Qy      2019 CAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTA 2078
             || ||||| ||  | || ||||||| | |  | || |  || ||   |    || ||
Db      6311 CAGGGCCACCGTCTACGCCTGTTTCAGAAGATCGCCCGCGCCTGTGCCAGCGTGGGGGCG 6370

Qy      2079 TCGCTGGTGATCGCCCATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTC 2138
               | ||   || ||||| || |  ||  | |   |||  ||| |  |   |||  |   |
Db      6371 GAGGTGACCATTGCCCACTGCGATGGCCTGACGCCCGCCCAGGCCGACTCGCTCTACGCC 6430

Qy      2139 GCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAG 2198
                |||||| ||  ||||      || ||||| ||||       | |     | ||  |
Db      6431 TGCGGCGCGACGGAGGTGGTCAGCTTTGTCGACCAGCCGCGCTACGTTGCCGAGGCTAAT 6490
```

```
Qy       2199 CTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCG 2258
Db       6491 CTGGTGATCACCCACGGCGGTCTCAATACCGTACTGGATGCGCTGGCTGCCGCGACGCCG 6550
Qy       2259 ATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCAT 2318
Db       6551 GTGCTGGCGGTGCCACTCTCTTTCGACCAGCCCGCCGTGGCTGCCCGGCTGGTCTATAAC 6610
Qy       2319 GACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAA 2378
Db       6611 GGGCTGGGTCGCCGGGTATCGCGCTTTGCCAGACAGCAGACGCTGGCGGATGAGATTGCC 6670
Qy       2379 CAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAG 2438
Db       6671 CAACTGCTGGGGGATGAGACGCTGCATCAGCGTCTGGCGACGGCCCGCCAGCAGCTTAAC 6730
Qy       2439 CGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAA 2498
Db       6731 GACGCCGGGGGCACGCCCCGTGCGGCGACCCTGATTGAACAGGCCAT------------- 6777
Qy       2499 GTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTG 2558
Db       6778 -----------AGCAGGGAGTGAGAGCGTATCGTGAGGGATCTGATTTTAGTCGGCGGCG 6826
Qy       2559 GACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCC 2618
Db       6827 GCCTGGCCAACGGGCTGATCGCCTGGCGTCTGCGCCAGCGCTACCCGCAGCTTAACCTGC 6886
Qy       2619 TGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCG 2678
Db       6887 TGCTGATCGAGGCCGGGGAGCAGCCCGGCGGGAACCATACCTGGTCATTCCATGAAGACG 6946
Qy       2679 ATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTT 2738
Db       6947 ATCTGACTCCCGGGCAGCACGCCTGGCTGGCCCCGCTGGTGGCCCACGCCTGGCCGGGCT 7006
Qy       2739 ATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCAT 2798
Db       7007 ATGAGGTGCAGTTTCCCGATCTTCGCCGTCGCCTCGCGCGCGGCTACTACTCCATTACCT 7066
Qy       2799 CAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAG 2858
Db       7067 CAGAGCGCTTTGCCGAGGCCCTGCATCAGGCGCTGGGGGAGAACATCTGGCTAAACTGTT 7126
Qy       2859 CCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGC 2918
Db       7127 CGGTGAGCGAGGTGTTACCCAATAGCGTGCGCCTTGCCAACGGTGAGGCGCTGCTTGCCG 7186
Qy       2919 AAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGG 2978
Db       7187 GAGCGGTGATTGACGGACGCGGCGTGACCGCCAGTTCGGCGATGCAAACCGGCTATCAGC 7246
Qy       2979 TGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGA 3038
Db       7247 TCTTTCTTGGTCAGCAGTGGCGGCTGACACAGCCCCACGGCCTGACCGTACCGATCCTGA 7306
Qy       3039 TGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCG 3098
Db       7307 TGGATGCCACGGTGGCGCAGCAGCAGGGCTATCGCTTTGTCTACACGCTGCCGCTCTCCG 7366
Qy       3099 CCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACA 3158
Db       7367 CCGACACGCTGCTGATCGAGGATACGCGCTACGCCAATGTCCCGCAGCGTGATGATAATG 7426
Qy       3159 CCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGC 3218
Db       7427 CCCTACGCCAGACGGTTACCGACTATGCTCACAGCAAAGGGTGGCAGCTGGCCCAGCTTG 7486
Qy       3219 TGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGC 3278
Db       7487 AACGCGAGGAGACCGGCTGTCTGCCGATTACCCTGGCGGGTGACATCCAGGCTCTGTGGG 7546
Qy       3279 AACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCG 3338

Db       7547 CCGATGCGCCGGGCGTGCCGCGCTCGGGAATGCGGGCTGGGCTATTTCACCCTACCACTG 7606
Qy       3339 GTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCG 3398
Db       7607 GCTATTCGCTGCCGCTGGCGGTGGCCCTTGCCGACGCGATTGCCGACAGCCCGCGGCTGG 7666
Qy       3399 ATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGC 3458
Db       7667 GCAGCGTTCCGCTCTATCAGCTCACCCGGCAGTTTGCCGAACGCCACTGGCGCAGGCAGG 7726
Qy       3459 GATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGC 3518
Db       7727 GATTCTTCCGCCTGCTGAACCGGATGCTTTTCCTGGCCGGGCGCGAGGAGAACCGCTGGC 7786
Qy       3519 GCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGC 3578
Db       7787 GGGTGATGCAGCGCTTTTATGGGCTGCCGGAGCCCACCGTAGAGCGCTTTTACGCCGGTC 7846
Qy       3579 AACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTG 3638
Db       7847 GGCTCTCTCTCTTTGATAAGGCCCGCATTTTGACGGGCAAGCCACCGGTTCCGCTGGGCG 7906
Qy       3639 AAGCGCTGCGCGCGCTGTTGA---ATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCAC 3695
Db       7907 AAGCCTGGCGGGCGGCGCTGAACCATTTTCCTGACAGACGAGATAAAGGATGAAAAAAAC 7966
Qy       3696 TTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGG 3755
Db       7967 CGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTCGCCTGCAGGCGGCAGG 8026
Qy       3756 CATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGA 3815
Db       8027 GATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTCGGGCCTACGTCTGGCA 8086
Qy       3816 GGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGA 3875
Db       8087 TGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCGATCCTACCGCGCTTGA 8146
Qy       3876 AGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAAC 3935
Db       8147 GGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCAGGCTGCTGCCGGTAAA 8206
Qy       3936 GCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCC 3995
Db       8207 ACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACTATGCTAACGACAGCGC 8266
Qy       3996 GCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATT 4055
Db       8267 CGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCGAGGGCTACCGGCGCTT 8326
Qy       4056 TCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTT 4115
Db       8327 TCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCCTCGGCAGCGTGCCGTT 8386
Qy       4116 TCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCG 4175
Db       8387 CCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTAAGCTCCAGGCGTGGCA 8446
Qy       4176 CAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTC 4235
Db       8447 GAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATCTGCGGCAGGCCTTCTC 8506
Qy       4236 CTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAAT 4295
Db       8507 GTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGTCCATCTACACCCTGAT 8566
Qy       4296 TCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGT 4355
Db       8567 CCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCGGCACCGGGGCGCTGGT 8626
Qy       4356 GCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGT 4415
```

```
Db      8627 GAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCGAACTCAACGCCCGGGT 8686

Qy      4416 GAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACG 4475
                  ||||||       ||  || ||| | |||   || | | | | ||  || ||
Db      8687 CGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGCTGGCGGATGGTCGGAT 8746

Qy      4476 CTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCT 4535
             ||| ||  ||| ||| || ||||| |   | ||||||||| | |||||  | || |||||
Db      8747 CTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACACCTATAAAAAGCTGCT 8806

Qy      4536 TCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAA 4595
               ||||||||||| |||    ||| || ||| || ||||| |||| ||| ||||||||||
Db      8807 CGGCCACCATCCGGTGGGGCAGAAGCGGGCGGCAGCGCTGGAGCGCAAGAGCATGAGCAA 8866

Qy      4596 CTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACAC 4655
             |||||||||||| ||||| || |||||||| ||||| |||   ||||| ||||||| ||
Db      8867 CTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCCAGCTGGCGCACCATAC 8926

Qy      4656 CGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCT 4715
             | |||||||||| || || || || ||| |||||||| ||||| || || | |||||  |||
Db      8927 CATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCTTTACCGGCAGCGCGCT 8986

Qy      4716 GGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACC 4775
             ||| || || || || || |||||||| ||||||||   | |||||| ||||| || ||
Db      8987 GGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCGATCCCTCGCTCGCGCC 9046

Qy      4776 GCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACAT 4835
              ||||||||||| |||||| || ||| | || ||||||||||||||| |||| |||     |
Db      9047 TCCCGGCTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATCTTGGCAACGCGCCGCT 9106

Qy      4836 CGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCA 4895
              ||||||   ||||| || |||    |||||| || || ||||  || || ||  |||
Db      9107 GGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTGACTACCTTGAAGAGCG 9166

Qy      4896 CTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTT 4955
             ||| ||||| ||  | |||   ||  | ||||| ||  | || ||||| |||   || ||
Db      9167 CTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCTTTACCCCGGCAGACTT 9226

Qy      4956 TCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCA 5015
              | |||||||||| |||| ||||  || ||||| || ||  | ||||||| |  |||| ||
Db      9227 CCACGACACGCTGGATGCGCATCTGGGATCGGCCTTCTCCATCGAGCCGCTGCTGACCCA 9286

Qy      5016 AAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGG 5075
             |||||||||||||||||||||| ||||||||   ||| ||   ||| |||||| ||||||||
Db      9287 AAGCGCCTGGTTCCGCCCGCACAACCGCGACAGCGACATTGCCAACCTCTACCTGGTGGG 9346

Qy      5076 TGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGC 5135
              || ||||| || || || ||||||| | || |||||  | |  ||||| || ||||||||
Db      9347 CGCAGGTACTCACCCTGGGGCGGGCATTCCTGGCGTAGTGGCCTCGGCGAAAGCCACCGC 9406

Qy      5136 CAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACG 5195
             ||| |||||| | |||||||      |||||  | || |||  | |||||| || |  |||
Db      9407 CAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCTGCTTGACCACGCCACG 9463

Qy      5196 CAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCG 5255
             || ||||||||    ||||||| ||||| |||||||| || |||||||| ||  |  |
Db      9464 CAGACCATGGCCAACGGCTCGAAAAGTTTTGCCACCGCTGCGAAGCTGTTCGACCCGGCC 9523

Qy      5256 ACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGAT 5315
             || ||||| |||  |||||||||| ||  | ||||| || |||||||||||| || ||||| 
Db      9524 ACCCGCCGTAGCGTGCTGATGCTCTACACCTGGTGCCGCCACTGCGATGACGTCATTGAC 9583

Qy      5316 GGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATG 5375
             |  || || |  |||   | | ||  ||     || |  || || ||       ||  ||
Db      9584 GACCAGACCCACGGCTTCGCCAGCGAGGCCGCGGCGGAGGAGGAGGCCACCCAGCGCCTG 9643

Qy      5376 CAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCG 5435
                |  ||||  |     ||||     || |     || ||  | ||| | || ||||| 
Db      9644 GCCCGGCTGCGCACGCTGACCCTGGCGGCGTTTGAAGGGGCCGAGATGCAGGATCCGGCC 9703

Qy      5436 TTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGAT 5495
             ||    || |||||||| |||||| | |  |||    |  |||  |   ||||| | |||
Db      9704 TTCGCTGCCTTTCAGGAGGTGGCGCTGACCCACGGTATTACGCCCCGCATGGCGCTCGAT 9763

Qy      5496 CATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACG 5555
             || || || ||||| || ||||| |      |    |  || |  | ||| || || |||
Db      9764 CACCTCGACGGCTTTGCGATGGACGTGGCTCAGACCCGCTATGTCACCTTTGAGGATACG 9823

Qy      5556 CTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGC 5615
             ||||| |||||||||||||| ||||||||||| ||| ||||||||||  | || ||||||
Db      9824 CTGCGCTACTGCTATCACGTGGCGGGCGTGGTGGGTCTGATGATGGCCAGGGTGATGGGC 9883

Qy      5616 GTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACT 5675
             ||||| || ||   ||||||| |||| ||||||||||| | || ||||| |||||||| || 
Db      9884 GTGCGGGATGAGCGGGTGCTGGATCGCGCCTGCGATCTGGGGCTGGCCTTCCAGCTGACG 9943

Qy      5676 AACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCC 5735
             || || || || || ||| | || ||||| |  | ||  |||||||||||||||
Db      9944 AATATCGCCCGGGATATTATTGACGATGCGGCTATTGACCGCTGCTATCTGCCCGCCGAG 10003

Qy      5736 TGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCG 5795
             |||||  |  | || ||  |   | | || |      | || | | |  |||| || |||
Db     10004 TGGCTGCAGGATGCCGGGCTGACCCCGGAGAACTATGCCGCGCGGGAGAATCGGGCCGCG 10063

Qy      5796 CTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGA 5855
             || ||     |||| | |||  |  | |  |  || || || || ||   |||  | |
Db     10064 CTGGCGCGGGTGGCGGAGCGGCTTATTGATGCCGCAGAGCCGTACTACATCTCCTCCCAG 10123

Qy      5856 TCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTT 5915
              ||||  |||  | | | ||| || |||||  || ||||| ||||| || || ||| |||| 
Db     10124 GCCGGGCTACACGATCTGCCGCCGCGCTGCGCCTGGGCGATCGCCACCGCCCGCAGCGTC 10183

Qy      5916 TATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAG 5975
             || || || || ||  | || ||  |    || || |    |||||||||  |   || |||
Db     10184 TACCGGGAGATCGGTATTAAGGTAAAAGCGGCGGGAGGCAGCGCCTGGGATCGCCGCCAG 10243

Qy      5976 CGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACT 6035
             | |||||| ||||||||||||| | ||  ||||| ||   |   | ||   ||  ||
Db     10244 CACACCAGCAAAGGTGAAAAAATTGCCATGCTGATGGCGGCACCGGGGCAGGTTATTCGG 10303

Qy      6036 TCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTT 6095
              ||     | |  |       ||||||||||| ||||| ||||||||||||   ||
Db     10304 GCGAAGACGACGAGGGTGACGCCGCGTCCGGCCGGTCTTTGGCAGCGTCCCGTTTAGGCG 10363

Qy      6096 TACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGA 6155
               || || ||||| |  |||||  | || || || |      ||    | ||||| |  |
Db     10364 GGCGGCCATGACGTTCACGCAGGATCGCCTGTAGGTCGGCAGGCTTGCGGGCGTAAATAA 10423

Qy      6156 AACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGT 6215
             |||| || || |||||||| || || ||||||||| || |||| |||| ||||| || | ||
Db     10424 AACCGAAGGAGACGCAGCCCTCCCGGCCGCGCACCGCGTGGTGCAGGCGGTGGGCGACGT 10483

Qy      6216 ATAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCA 6275
             | |  ||||| || |||||   ||||||||||  || ||||| |||||||| ||||||||||
Db     10484 AGAGCCGCTTCAGGTAGCCCCGGCGCGGGATCCAGTGGAAGGGCCAGCGCTGATGCACCA 10543

Qy      6276 GGCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACT 6335
             | || ||||||||||| ||||||||||  ||| ||   |||||| |||   ||||||||||
Db     10544 GACCGTCGTGCACCAGGAAGTAGAGCAGGCCATAGACCGTCATGCCGCAGCCAATCCACT 10603

Qy      6336 GCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACA 6395
             |||| ||||| | |||   |   || ||   ||||||| | || ||||   | |||||||
Db     10604 GCAGGGGCCAAACGCCCGCCGTGCCCACGGCAATCAGCGCGATAGCCACCCCGGCAAACA 10663

Qy      6396 CCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGAT 6455
             |||||||| | |||||||| ||||||||    ||  || ||||   |||||| |||  ||
Db     10664 CCACCGCAAAGAGATCGTTTAGCTCAAATACGCCCTTGCGCGGGGTATGGTGTGACTCAT 10723

Qy      6456 GCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCA 6515
             |||||| |||||||||| |||||||| |||||    || |  | ||| || | |   ||||
Db     10724 GCCAGCGCCATCCCCAGCCGTGCATAATGTAGCGGTGGGTAAACGCGGCGATGCCCTCCA 10783

Qy      6516 TCACCACCACGGTTGCCAACAAGATAA 6542
             || | |  ||| | |    |  | ||| |
Db     10784 TCGCAATAACGCTCAAGATGACGATTA 10810
```

```
RESULT 7
ADP74122
ID   ADP74122 standard; DNA; 4624 BP.
XX
AC   ADP74122;
XX
DT   26-AUG-2004  (first entry)
XX
DE   Pantoea ananatis DNA fragment encoding CrtY, Crt1 and crt8.
XX
KW   ketocarotenoid; ketolase; carotenoid; hydroxylase; beta-cyclase;
KW   astaxanthin; adonirubin; canthaxanthin; adoxanthin; zeaxanthin;
KW   antioxidant; pigment; animal feed; trout; salmon; shrimp;
KW   food supplement; gene; ds.
XX
OS   Pantoea ananatis.
XX
FH   Key             Location/Qualifiers
FT   CDS             128..1267
FT                   /*tag=  a
FT                   /product= "CrtY"
FT   CDS             1288..2766
FT                   /*tag=  b
FT                   /product= "Crt1"
FT   CDS             2802..3692
FT                   /*tag=  c
FT                   /product= "crt8"
XX
PN   DE10253112-A1.
XX
PD   03-JUN-2004.
XX
PF   13-NOV-2002; 2002DE-01053112.
XX
PR   13-NOV-2002; 2002DE-01053112.
XX
PA   (SUNG-) SUNGENE GMBH & CO KGAA.
XX
PI   Sauer M,  Flachmann R,  Klebsattel M,  Schopfer CR;
XX
DR   WPI; 2004-489014/47.
DR   P-PSDB; ADP74123, ADP74124, ADP74125.
XX
PT   Production of ketocarotenoids with low hydroxylated by-product content,
PT   for use e.g. in pigmenting feedstuffs, by culturing genetically modified
PT   organisms having modified ketolase activity.
XX
PS   Example 2; SEQ ID NO 24; 101pp; German.
XX
CC   This invention describes a novel method for the production of
CC   ketocarotenoids which involves culturing genetically modified organisms
CC   having modified ketolase (KLA) activity. The starting microorganisms
CC   produce carotenoids (naturally or by genetic supplementation), and are
CC   specifically microorganisms (especially bacteria, yeasts, algae or fungi)
CC   or plants. Microorganisms and plants involved in the method could include
CC   Escherichia, Flavobacterium, Nostoc, Syneochocystis, Hansenula, Fusarium,
CC   Dunaliella, Ranunculaceae, Cannabaceae, Brassiceae, Amaranthaceae,
CC   Solanaceae, Lamiaceae, Acacia, Calendula, Gentiana, Helianthus, Linum,
CC   Rhododendron, Spartium and Zinnia. The modified microorganisms
CC   additionally show elevated hydroxylase and/or beta-cyclase activity. The
CC   ketocarotenoids described in the invention include astaxanthin,
CC   adonirubin, canthaxanthin, adoxanthin and zeaxanthin. Ketocarotenoids are
CC   natural antioxidants and pigments, especially useful as pigmenting
CC   additives in animal feed, specifically feed for trout, salmon or shrimps.
CC   Ketocarotenoid-producing genetically modified organisms can be used as
CC   feedstuffs or foodstuffs, in the production of ketocarotenoid-containing
CC   extracts or for producing feed or food supplements. The process provides
CC   large amounts of ketocarotenoids having a low content of hydroxylated by-
CC   products, especially in the case of astaxanthin, preferably due to
CC   expression of at least one nucleic acid encoding hydroxylase and/or beta-
CC   cyclase. DNA encoding the whole primary ketolase sequence from Nostoc sp.
CC   strain PCC7120 was isolated, amplified by PCR and used to produce a
CC   plasmid pNOSTF-G. A plasmid pMCL-Crt-YIBZ/idi/gps, for the synthesis of
CC   zeaxanthin in Escherichia coli, was constructed in 3 stages via the
CC   intermediate stages pMCL-CrtYIBZ and pMCL-CrtYIBX/idi, using the high
CC   copy number plasmid vector pMCL200. Escherichia coli strain TOP10 was
CC   transformed with the plasmids pNOSTF-G and pMCL-Crt-YIBZ/idi/gps to give
CC   carotenoid producing strain.
XX
SQ   Sequence 4624 BP; 1026 A; 1243 C; 1270 G; 1085 T; 0 U; 0 Other;

  Query Match             27.2%;  Score 1900.6;  DB 12;  Length 4624;
  Best Local Similarity   66.5%;  Pred. No. 0;
  Matches 2787;  Conservative    0;  Mismatches 1389;  Indels   13;  Gaps    4;

Qy       2405 GCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGC 2464
              ||  || ||| |    || ||| |    || |     ||||| ||   |  |   || ||
Db         12 GCAGCGCATGGCGAAAATCCAGACAGCCCTTCGTTTGGCAGGGGGCACCATGGCCGCTGC 71

Qy       2465 CGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATG 2524
              ||| ||| | |||||||   |||| || | |  || |   ||   |  || |     |||
Db         72 CGATATCATTGAGCAGGTTATGTG-CACCGGTCAGCCTGTCTTAAGTGGGAGCGGCTATG 130

Qy       2525 CGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTG 2584
              |   |||| || ||| ||||| | || || |||||||| ||||||||| | || || |||
Db        131 CAACCGCATTATGATCTGATTCTCGTGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTG 190

Qy       2585 CGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCG 2644
              ||||| |  ||   ||| ||  |  ||       || |  | ||   ||   | ||  ||
Db        191 CGTCTTCAGCAGCAGCAACCTGATATGCGTATTTTGCTTATCGACGCCGCACCCCAGGCG 250

Qy       2645 GCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGG 2704
              |  || ||||||||| ||||| ||||| |||   ||| | |  |     |||| ||| |||
Db        251 GGCGGGAATCATACGTGGTCATTTCACCACGATGATTTGACTGAGAGCCAACATCGTTGG 310

Qy       2705 CTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGC 2764
               |    |||||| |       |  ||| | |  ||||||||| || |||||  | |  ||
Db        311 ATAGCTCCGCTGGTGGTTCATCACTGGCCCGACTATCAGGTACGCTTTCCCACACGCCGT 370

Qy       2765 CGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTAC 2824
              || || ||| || | |  || | ||  ||  | ||      ||| ||      | |  |
Db        371 CGTAAGCTGAACAGCGGCTACTTTTGTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAG 430

Qy       2825 GCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAG 2884
                   | | ||     |  ||||||   | || || ||   | |||| ||  |
Db        431 CGACAGTTTGGCCCGCACTTGTGGATGGATACCGCGGTCGCAGAGGTTAATGCGGAATCT 490

Qy       2885 GTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTG 2944
              ||   | ||    | || |  |   | | ||| |  | ||||||||| || || ||
Db        491 GTTCGGTTGAAAAAGGGTCAGGTTATCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTAT 550

Qy       2945 CAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTG 3004
                | | |   ||   |||    |||| |  |||| |||| |||| || || ||||   ||
Db        551 GCGGCAAATTCAGCACTGAGCGTGGGCTTCCAGGCGTTTATTGGCCAGGAATGGCGATTG 610

Qy       3005 GCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCG 3064
                 || ||||| ||  |   |   || ||  | ||||||||||| ||||||||||||
Db        611 AGCCACCCGCATGGTTTATCGTCTCCCATTATCATGGATGCCACGGTCGATCAGCAAAAT 670

Qy       3065 GGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACC 3124
              |||||||| || || ||||  |||||||||    |     |  | || |||||||| ||
Db        671 GGTTATCGCTTCGTGTACAGCCTGCCGCTCTCGCCGACCAGATTGTTAATTGAAGACACG 730

Qy       3125 CATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTAT 3184
              || ||  || |  |  |  |  | |      |   ||| || ||  | ||   |||||||
Db        731 CACTATATTGATAATGCGACATTAGATCCTGAATGCGCGCGGCAAAATATTTGCGACTAT 790

Qy       3185 GCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCG 3244
              |||   || || || |||  |||    || |||||||| ||||| || |||   |||||
Db        791 GCCGCGCAACAGGGTTGGCAGCTTCAGACACTGCTGCGAGAAGAACAGGGCGCCTTACCC 850

Qy       3245 ATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGC 3304
              ||||| |||   |||||   |||   |||||||   ||||||||||  ||  || || ||
Db        851 ATTACTCTGTCGGGCAATGCCGACGCATTCTGG---CAGCAGCGCCCCCTGGCCTGTAGT 907
```

```
Qy         3305 GGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCG 3364
                ||  | || ||||| ||||| ||| | |||||||| || ||  |||||   || || ||
Db          908 GGATTACGTGCCGGTCTGTTCCATCCTACCACCGGCTATTCACTGCCGCTGGCGGTTGCC 967

Qy         3365 CTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATC 3424
                 | || ||     |    || ||      |       |   | |  |   |||    ||
Db          968 GTGGCCGACCGCCTGAGTGCACTTGATGTCTTTACGTCGGCCTCAATTCACCATGCCATT 1027

Qy         3425 GAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATG 3484
                   |  |||||||| ||||  ||||   | ||| | ||||||||  ||||| || ||||||
Db         1028 ACGCATTTTGCCCGCGAGCGCTGGCAGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATG 1087

Qy         3485 CTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTC 3544
                |||||||| |||||     |  |    |||||||| || ||||| ||||||||  |  |
Db         1088 CTGTTTTTAGCCGGACCCGCCGATTCACGCTGGCGGGTTATGCAGCGTTTTTATGGTTTA 1147

Qy         3545 GATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGG 3604
                  ||  |  ||||||  ||| ||||| || ||  ||||     ||  ||||      ||
Db         1148 CCTGAAGATTTAATTGCCCGTTTTTATGCGGGAAAACTCACGCTGACCGATCGGCTACGT 1207

Qy         3605 ATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCT 3664
                |||||| ||||||||||||| || ||  |    | |||  |||  ||  |  |||   ||
Db         1208 ATTCTGAGCGGCAAGCCGCCTGTTCCGGTATTAGCAGCATTGCAAGCCATTATGACGACT 1267

Qy         3665 GTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCC 3724
                        | |   ||  | |||||||  |||   || ||||| |||||||| || ||||
Db         1268 CATCGTTAAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCC 1327

Qy         3725 TGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCG 3784
                |||| ||||| ||||| || ||||| ||||| || ||   |    | || || || || |
Db         1328 TGGCACTGGCAATTCGTCTACAAGCTGCGGGGATCCCCGTCTTACTGCTTGAACAACGTG 1387

Qy         3785 ACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGAC 3844
                | ||||| ||||| || || ||||| |  |||||    || |||||||| ||||| || |
Db         1388 ATAAACCCGGCGGTCGGGCTTATGTCTACGAGGATCAGGGGTTTACCTTTGATGCAGGCC 1447

Qy         3845 CCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGC 3904
                | ||||| |||||||||||||| ||||| |||||  ||||  | ||||||||||||  |
Db         1448 CGACGGTTATCACCGATCCCAGTGCCATTGAAGAACTGTTTGCACTGGCAGGAAAACAGT 1507

Qy         3905 TCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATG 3964
                | |  || || ||||| ||| ||||||| |||| || || |||||||| |||||    |
Db         1508 TAAAAGAGTATGTCGAACTGCTGCCGGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAG 1567

Qy         3965 GCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCA 4024
                | ||     || ||||||| ||  ||||  | | ||| ||   ||||||      ||| |
Db         1568 GGAAGGTCTTTAATTACGATAACGATCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTA 1627

Qy         4025 ATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAG 4084
                |||| |  ||||| ||||| |||||||| ||||| | ||||||||| |  || |||| ||
Db         1628 ATCCCCGCGATGTCGAAGGTTATCGTCAGTTTCTGGACTATTCACGCGCGGTGTTTAAAG 1687

Qy         4085 AGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCG 4144
                | || ||||| || ||||| || || || ||| |   | |  | |||||||| |||| ||
Db         1688 AAGGCTATCTAAAGCTCGGTACTGTCCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCG 1747

Qy         4145 CGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTA 4204
                | || ||  |||     |||||| |||||| | |   | ||||| | ||| || |  || |
Db         1748 CACCTCAACTGGCGAAACTGCAGGCATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACA 1807

Qy         4205 TTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATC 4264
                |  | || || |||||||| ||||||||||| ||||||||  || |||||||||| ||||
Db         1808 TCGAAGATGAACATCTGCGCCAGGCGTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATC 1867

Qy         4265 CTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGT 4324
                | || || || ||||| || ||||| || || || |||||||||||||| |||||||| |
Db         1868 CCTTCGCCACCTCATCCATTTATACGTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCT 1927

Qy         4325 GGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACT 4384
                |||||||||| ||||||||||||||  | || ||||| |||      |||||  ||||
Db         1928 GGTTTCCGCGTGGCGGCACCGGCGCATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATC 1987

Qy         4385 TGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCA 4444
```

```
                |||| |||||  |  |  | || |||  ||| |||||  ||||||| |  || ||    |
Db         1988 TGGGTGGCGAAGTCGTGTTAAACGCCAGAGTCAGCCATATGGAAACGACAGGAAACAAGA 2047

Qy         4445 TTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATG 4504
                ||   | ||| || ||||||| ||| ||  | |||    |    || || || || ||||
Db         2048 TTGAAGCCGTGCATTTAGAGGACGGTCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATG 2107

Qy         4505 CCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTG 4564
                | || ||||| ||||||||   | | ||| |  |||| || ||    ||  | || |
Db         2108 CAGATGTGGTTCATACCTATCGCGACCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGT 2167

Qy         4565 CGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGA 4624
                | |     ||| ||  |||||||||||| ||||| |||||||| ||||||||||||  |||
Db         2168 CCAACAAACTGCAGACTAAGCGCATGAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGA 2227

Qy         4625 ATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGT 4684
                |||| |  ||||| ||||||||||| ||||| || ||||| ||||||||||| || |||
Db         2228 ATCACCATCATGATCAGCTCGCGCATCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGC 2287

Qy         4685 TGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACG 4744
                |||| || || |||| ||           || ||||| || || |||||||| |||||||
Db         2288 TGATTGACGAAATTTTTAATCATGATGGCCTCGCAGAGGACTTCTCACTTTATCTGCACG 2347

Qy         4745 CGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAG 4804
                |||||||   ||  ||| |||| ||||| ||    || ||||||||| |  |||||||| |
Db         2348 CGCCCTGTGTCACGGATTCGTCACTGGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGG 2407

Qy         4805 CGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGC 4864
                |||||||||||||| | ||||||||  || ||||||||      || || ||     | |
Db         2408 CGCCGGTGCCGCATTTAGGCACCGCGAACCTCGACTGGACGGTTGAGGGGCCAAAACTAC 2467

Qy         4865 GCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAG 4924
                |||| || |||||||| || || ||||||||| |||||||| || |||||    ||  | |
Db         2468 GCGACCGTATTTTTGCGTACCTTGAGCAGCATTACATGCCTGGCTTACGGAGTCAGCTGG 2527

Qy         4925 TGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCT 4984
                | || ||| | |||||||||||||||||||||||||||  |||  ||||| |||| ||||
Db         2528 TCACGCACCGGATGTTTACGCCGTTTGATTTTCGCGACCAGCTTAATGCCTATCATGGCT 2587

Qy         4985 CGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCG 5044
                | || |||||  |||||||  || | || || ||||||||||| || |||||||||||||
Db         2588 CAGCCTTTTCTGTGGAGCCCGTTCTTACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCG 2647

Qy         5045 ATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGC 5104
                ||      || |  |||||||| ||||| || || || ||||||| |||||| ||| | |
Db         2648 ATAAAACCATTACTAATCTCTACCTGGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTC 2707

Qy         5105 CCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAAT 5164
                | ||||| ||||| ||||| || || || ||  |  |||||||||||||||        |
Db         2708 CTGGCGTCATCGGCTCGGCAAAAGCGACAGCAGGTTTGATGCTGGAGGATC---TGATTT 2764

Qy         5165 GAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTT 5224
                ||||   | | | |||||  | || |      |||| ||||| || ||||||||| |||||
Db         2765 GAATAATCCGTCGTTACTCAATCATGCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTT 2824

Qy         5225 CGCCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGC 5284
                 || || ||| | ||| | ||||||||   || || ||||||   |||||||| || ||
Db         2825 TGCGACAGCCTCAAAGTTATTTGATGCAAAAACCCGGCGCAGCGTACTGATGCTCTACGC 2884

Qy         5285 GTGGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCA 5344
                 ||||| || || || || ||||| ||||| |  || |||||||||      | |  |||
Db         2885 CTGGTGCCGCCATTGTGACGATGTTATTGACGATCAGACGCTGGGCTTTCAGGCCCGGCA 2944

Qy         5345 GCATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGC 5404
                || |||| |  ||    |  |   |||| ||  ||| ||  | ||  |||| ||||  ||
Db         2945 GCCTGCCTTACAAACGCCCGAACAACGTCTGATGCAACTTGAGATGAAAACGCGCCAGGC 3004

Qy         5405 CTACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCAT 5464
                |||    ||  |||| ||| | ||||||||||||  ||| ||||||||||||||| ||
Db         3005 CTATGCAGGATCGCAGATGCACGAACCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGC 3064

Qy         5465 TCACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACG 5524
                |||  |  |  | |        |||||||||||||||||||||||||| ||||||| |||
Db       3065 TCATGATATCGCCCCGGCTTACGCGTTTGATCATCTGGAAGGCTTCGCCATGGATGTACG 3124

Qy       5525 CAACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGT 5584
              | | |  || |||       | ||||| |||||||| || ||||||||||| || |||||
Db       3125 CGAAGCGCAATACAGCCAACTGGATGATACGCTGCGCTATTGCTATCACGTTGCAGGCGT 3184

Qy       5585 GGTCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGC 5644
               ||||| |||||||||||||   | ||||||||||| ||  | ||   ||| || | |||
Db       3185 TGTCGGCTTGATGATGGCGCAAATCATGGGCGTGCGGGATAACGCCACGCTGGACCGCGC 3244

Qy       5645 CTGCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGC 5704
              ||| ||  | || ||||| || ||| | || || ||||| ||||| |||| || |||||
Db       3245 CTGTGACCTTGGGCTGGCATTTCAGTTGACCAATATTGCTCGCGATATTGTGGACGATGC 3304

Qy       5705 CGAAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGA 5764
                |    || ||||| |||||||||  |  |||||| || || |  ||  ||   |   ||
Db       3305 GCATGCGGGCCGCTGTTATCTGCCGGCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGA 3364

Qy       5765 TACGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGC 5824
               |      | |||||  || | |||   ||||||   |    | ||    |||||  |||
Db       3365 GAATTATGCGGCACCTGAAAACCGTCAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCA 3424

Qy       5825 GGAGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTC 5884
              ||| || ||||| || |||   || ||   | ||||  |  | || || || ||||| ||
Db       3425 GGAAGCAGAACCTTACTATTTGTCTGCCACAGCCGGCCTGGCAGGGTTGCCCCTGCGTTC 3484

Qy       5885 GGCGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCA 5944
               || ||||| |||||||||||       |||||| ||  |||| || |||||||||| | ||
Db       3485 CGCCTGGGCAATCGCTACGGCGAAGCAGGTTTACCGGAAAATAGGTGTCAAAGTTGAACA 3544

Qy       5945 CGCCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCT 6004
               ||||||  ||| |||||||||   ||||||   || |  |     |||||| |  ||||
Db       3545 GGCCGGTCAGCAAGCCTGGGATCAGCGGCAGTCAACGACCACGCCCGAAAAATTAACGCT 3604

Qy       6005 GCTGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCC 6064
              |||| ||   |   | |||  |||  | ||||| ||  ||     || |   || || ||
Db       3605 GCTGCTGGCCGCCTCTGGTCAGGCCCTTACTTCCCGGATGCGGGCTCATCCTCCCCGCCC 3664

Qy       6065 GGCTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCT 6124
               ||   ||| |||||||| || |      | ||   || || |  |  || ||||| ||
Db       3665 TGCGCATCTCTGGCAGCGCCCGC------TCTAGCGCCATGTCTTTCCCGGAGCGTCGCC 3718

Qy       6125 TGCAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCG 6184
              || || ||   ||| || ||||| |||||||| ||||| || || || |||||    ||
Db       3719 TGAAGTTTTGACAGGGGCGGCGCATAGAGGAAGCCAAAAGAAACACAACCTTCTTTGCCC 3778

Qy       6185 CGCACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGG 6244
              |  || || ||||||||| ||||||||||| || || || || || ||||| ||||| ||
Db       3779 CTGACGGCGTGATGCATACGGTGCGCCATATACAACCGTTTGAGGTAGCCCTTGCGTGGA 3838

Qy       6245 ATATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCG 6304
              ||||||||||| ||||| |||||||||||||| || |||||||||||| ||  |||||
Db       3839 ATATAGCGGAATGGCCAACGTTGATGCACCAGCCCGTCGTGCACCATAAAATAGAGTAAT 3898

Qy       6305 CCGTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACA 6364
              || |||| |||||| || || ||||||||||| ||||||||||||| |||      || | |
Db       3899 CCATACGCCGTCATACCTGCGCCAATCCACTGGAGCGGCCACATTCCTGTACTGCCCAGA 3958

Qy       6365 TAAATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAAC 6424
              ||||||||||  ||||  | | | |||||  ||||| |||||||||||||||| |  ||||||
Db       3959 TAAATCAGCAGGATCGATAATGCAGCAAAAACCACGGCATAAAGATCGTTAACTTCAAAC 4018

Qy       6425 TTACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATG 6484
                |||  |  || |||||||||| || || ||||||||| ||||| ||||| ||||||||||||
Db       4019 GCACCTTTACGCGGTTCATGATGTGAAAGATGCCATCCCCAACCCCAGCCGTGCATGATG 4078

Qy       6485 TATTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGC 6544
              ||||| || | ||| || || |  | ||||||  | | ||||||  | || | |||  ||
Db       4079 TATTTGTGTGCCAGTGCAGCAATCACTTCCATGCCAATCACGGTAACGAAAACGATCAGG 4138

Qy       6545 ACGTTCCATAACCAGAGCATTGTTCGTCCATTTGTGGAAAAGGGAAGTA 6593
               | ||||| | ||| | |||  ||  |||  | | |      || || |
Db       4139 GCATTCCAAATCCACAACATAATTTCTCCGGTAGAGACGTCTGGCAGCA 4187
```

```
RESULT 8
ABZ69177
ID   ABZ69177 standard; DNA; 7494 BP.
XX
AC   ABZ69177;
XX
DT   11-AUG-2003  (first entry)
XX
DE   Vector construct pBScaro14 coding sequence.
XX
KW   Biosynthesis; enzyme; antibacterial; protozoacide; herbicide; isoprenoid;
KW   inhibitor; immunostimulant; gene; ds.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   DE10201458-A1.
XX
PD   17-OCT-2002.
XX
PF   16-JAN-2002; 2002DE-01001458.
XX
PR   11-APR-2001; 2001DE-01018166.
PR   22-JUN-2001; 2001DE-01030236.
PR   09-NOV-2001; 2001DE-01055084.
XX
PA   (BACH/) BACHER A.
PA   (ROHD/) ROHDICH F.
XX
PI   Adam P,  Amsingler S,  Bacher A,  Eisenreich W,  Hecht S,  Rohdich F;
XX
DR   WPI; 2003-185556/19.
XX
PT   New proteins involved in isoprenoid biosynthesis, useful in screening for
PT   inhibitors, also new intermediates, potential therapeutic agents, nucleic
PT   acids and antibodies.
XX
PS   Disclosure; Page 45-47; 78pp; German.
XX
CC   The present invention relates to proteins capable of the enzymatic
CC   conversion of 2C-methyl-D-erythritol-2,4-cyclodiphosphate to 1-hydroxy-2-
CC   methyl-2-butenyl-4-diphosphate. The proteins are useful in screening a
CC   chemical library for inhibitors of isoprenoid biosynthesis. Cells and
CC   organisms that express them are useful for in vivo/in vitro production of
CC   isotopically labeled biosynthetic intermediates and products of the
CC   mevalonate-independent pathway of isoprenoid synthesis. Such
CC   intermediates and products are used to screen for genes, enzymes or
CC   inhibitors of isoprenoid/terpenoid biosynthesis, potentially useful as
CC   antibiotics, antimalarials and herbicides and as immunostimulators,
CC   particularly for treating bacterial infection. Antibodies against the
CC   intermediates are useful for detecting pathogens, in body fluids. The
CC   present sequence is a coding sequence used in the exemplification of the
CC   invention
XX
SQ   Sequence 7494 BP; 1722 A; 1935 C; 2026 G; 1811 T; 0 U; 0 Other;

  Query Match             23.1%;  Score 1618.8;  DB 8;  Length 7494;
  Best Local Similarity   66.0%;  Pred. No. 0;
  Matches 2375;  Conservative    0;  Mismatches 1217;  Indels    6;  Gaps    2;

Qy       2490 CAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGG 2549
              || | |  || |   ||   |  || |     ||||   |||| || ||| ||||| | |
Db       3213 CACCGGTCAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCG 3272

Qy       2550 TCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAAC 2609
              | || |||||||| |||||||||| | || || ||||||||| |  ||   ||| ||  |
Db       3273 TGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATA 3332

Qy       2610 TGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTC 2669
              ||       || |  | ||   ||   | ||  |||  || |||||||| ||||| ||||
Db       3333 TGCGTATTTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTC 3392
```

```
Qy   2670 ATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTT 2729
Db   3393 ACCACGATGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACT 3452

Qy   2730 GGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTT 2789
Db   3453 GGCCCGACTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTT 3512

Qy   2790 CCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGA 2849
Db   3513 GTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGA 3572

Qy   2850 CAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAAC 2909
Db   3573 TGGATACCGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTA 3632

Qy   2910 TTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGG 2969
Db   3633 TCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGG 3692

Qy   2970 GTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGC 3029
Db   3693 GCTTCCAGGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTC 3752

Qy   3030 CGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGC 3089
Db   3753 CCATTATCATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGC 3812

Qy   3090 CGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGG 3149
Db   3813 CGCTCTCGCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAG 3872

Qy   3150 CGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGA 3209
Db   3873 ATCCTGAATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTC 3932

Qy   3210 GTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATC 3269
Db   3933 AGACACTGCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACG 3992

Qy   3270 GATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATG 3329
Db   3993 CATTCTGG---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATC 4049

Qy   3330 CCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGT 3389
Db   4050 CTACCACCGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTG 4109

Qy   3390 TGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGC 3449
Db   4110 ATGTCTTTACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGC 4169

Qy   3450 GCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGC 3509
Db   4170 AGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATT 4229

Qy   3510 AGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTT 3569
Db   4230 CACGCTGGCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTT 4289

Qy   3570 ACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGC 3629
Db   4290 ATGCGGGAAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTC 4349

Qy   3630 CCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAA 3689
Db   4350 CGGTATTAGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAA 4409

Qy   3690 ACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGC 3749
Db   4410 ACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGC 4469
```

```
Qy   3750 GGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGT 3809
Db   4470 TGCGGGGATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGT 4529

Qy   3810 GTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGC 3869
Db   4530 CTACGAGGATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGC 4589

Qy   3870 CATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCC 3929
Db   4590 CATTGAAGAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCC 4649

Qy   3930 GGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAA 3989
Db   4650 GGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGA 4709

Qy   3990 TCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCG 4049
Db   4710 TCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCG 4769

Qy   4050 TCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGT 4109
Db   4770 TCAGTTTCTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGT 4829

Qy   4110 GCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGC 4169
Db   4830 CCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGC 4889

Qy   4170 ATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGC 4229
Db   4890 ATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGC 4949

Qy   4230 GTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATAC 4289
Db   4950 GTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATAC 5009

Qy   4290 CTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGC 4349
Db   5010 GTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGC 5069

Qy   4350 GCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGC 4409
Db   5070 ATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGC 5129

Qy   4410 CGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGG 4469
Db   5130 CAGAGTCAGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGG 5189

Qy   4470 ACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAA 4529
Db   5190 TCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGA 5249

Qy   4530 ACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCAT 4589
Db   5250 CCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCAT 5309

Qy   4590 GAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCA 4649
Db   5310 GAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCA 5369

Qy   4650 CCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAG 4709
Db   5370 TCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGA 5429

Qy   4710 CCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCT 4769
Db   5430 TGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACT 5489

Qy   4770 GGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGC 4829
Db   5490 GGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGC 5549

Qy   4830 TGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGA 4889
```

```
               || ||||||||      || || ||     | ||||| || |||||||| || || ||
Db      5550 GAACCTCGACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTGCGTACCTTGA 5609

Qy      4890 GCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTT 4949
             |||||| |||||||| || |||||    ||  | || || ||| | ||||||||||||||
Db      5610 GCAGCATTACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTT 5669

Qy      4950 TGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTT 5009
             ||||||||||||||  |||  ||||| |||| ||||| || |||||  |||||||  || |
Db      5670 TGATTTTCGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCT 5729

Qy      5010 GACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCT 5069
              || || |||||||||| || |||||||||||||||||   || |  |||||||| ||
Db      5730 TACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCT 5789

Qy      5070 GGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGC 5129
             ||| || || || ||||||||| ||||| ||| | || ||||| ||||| |||| || ||
Db      5790 GGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGC 5849

Qy      5130 CACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAA 5189
              || ||  |  |||||||||||||||        |||||   | | | ||||||  | || 
Db      5850 GACAGCAGGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCAT 5906

Qy      5190 GTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGAT 5249
             |      |||| ||||| || |||||||| ||||| || || ||| | ||| | ||||||
Db      5907 GCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGAT 5966

Qy      5250 GCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTG 5309
             |||   || || ||||||   |||||||| || || ||||| || || || ||| ||||| 
Db      5967 GCAAAAACCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTT 6026

Qy      5310 ATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCA 5369
             ||||| |  || |||||||||      | |  ||||| |||| |  ||    |  |   |
Db      6027 ATTGACGATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAA 6086

Qy      5370 CGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAA 5429
             ||| ||  ||| ||  | ||  |||| ||||  |||||    ||  |||| ||| | |||
Db      6087 CGTCTGATGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAA 6146

Qy      5430 CCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCG 5489
             |||||||||  ||| ||||||||||||||| ||   |||  |  |  | |        |||
Db      6147 CCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCG 6206

Qy      5490 TTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGAT 5549
             ||||||||||||||||||||||| ||||||| |||| | |  || |||       | |||
Db      6207 TTTGATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGAT 6266

Qy      5550 GACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTA 5609
             || ||||||||| || ||||||||||| || ||||| ||||| ||||||||||||||   | 
Db      6267 GATACGCTGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATC 6326

Qy      5610 ATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAG 5669
             ||||||||||| ||  | ||   ||| || | |||||| ||  | || ||||| || |||
Db      6327 ATGGGCGTGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAG 6386

Qy      5670 CTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAATGGTCGCTGCTATCTGCCG 5729
              | || || ||||| ||||| ||||| || |||||  |    || ||||| |||||||||
Db      6387 TTGACCAATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCG 6446

Qy      5730 CAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGT 5789
               |  |||||| || || |  ||  |   |   || |      | |||||  || | |||
Db      6447 GCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGT 6506

Qy      5790 GCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCG 5849
                ||||   |    | ||    ||||| |||  ||| || ||||| ||  |||   || 
Db      6507 CAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCT 6566

Qy      5850 GCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGC 5909
             ||   | ||||  |  | || || || ||| || || || |||||| ||||||||||| 
Db      6567 GCCACAGCCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAG 6626

Qy      5910 GGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCA 5969
                ||||| ||  |||| || ||||||||| | || ||||||  ||| |||||||||||   
Db      6627 CAGGTTTACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAG 6686

Qy      5970 CGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCG 6029
             ||||||   || |  |     |||||| |  |||||||| ||   |   | |||  ||| 
Db      6687 CGGCAGTCAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCC 6746

Qy      6030 ATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTC 6087
              | ||||| ||  ||     || |   || || || ||   ||| |||||||| || |
Db      6747 CTTACTTCCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC 6804


RESULT 9
ABZ69178
ID   ABZ69178 standard; DNA; 8547 BP.
XX
AC   ABZ69178;
XX
DT   11-AUG-2003  (first entry)
XX
DE   Vector construct pACYCcaro14 coding sequence.
XX
KW   Biosynthesis; enzyme; antibacterial; protozoacide; herbicide; isoprenoid;
KW   inhibitor; immunostimulant; gene; ds.
XX
OS   Unidentified.
OS   Synthetic.
XX
PN   DE10201458-A1.
XX
PD   17-OCT-2002.
XX
PF   16-JAN-2002; 2002DE-01001458.
XX
PR   11-APR-2001; 2001DE-01018166.
PR   22-JUN-2001; 2001DE-01030236.
PR   09-NOV-2001; 2001DE-01055084.
XX
PA   (BACH/) BACHER A.
PA   (ROHD/) ROHDICH F.
XX
PI   Adam P,  Amsingler S,  Bacher A,  Eisenreich W,  Hecht S,  Rohdich F;
XX
DR   WPI; 2003-185556/19.
XX
PT   New proteins involved in isoprenoid biosynthesis, useful in screening for
PT   inhibitors, also new intermediates, potential therapeutic agents, nucleic
PT   acids and antibodies.
XX
PS   Example 7; Page 48-50; 78pp; German.
XX
CC   The present invention relates to proteins capable of the enzymatic
CC   conversion of 2C-methyl-D-erythritol-2,4-cyclodiphosphate to 1-hydroxy-2-
CC   methyl-2-butenyl-4-diphosphate. The proteins are useful in screening a
CC   chemical library for inhibitors of isoprenoid biosynthesis. Cells and
CC   organisms that express them are useful for in vivo/in vitro production of
CC   isotopically labeled biosynthetic intermediates and products of the
CC   mevalonate-independent pathway of isoprenoid synthesis. Such
CC   intermediates and products are used to screen for genes, enzymes or
CC   inhibitors of isoprenoid/terpenoid biosynthesis, potentially useful as
CC   antibiotics, antimalarials and herbicides and as immunostimulators,
CC   particularly for treating bacterial infection. Antibodies against the
CC   intermediates are useful for detecting pathogens, in body fluids. The
CC   present sequence is a coding sequence used in the exemplification of the
CC   invention
XX
SQ   Sequence 8547 BP; 1884 A; 2296 C; 2279 G; 2088 T; 0 U; 0 Other;

  Query Match             23.1%;  Score 1618.8;  DB 8;  Length 8547;
  Best Local Similarity   66.0%;  Pred. No. 0;
  Matches 2375;  Conservative    0;  Mismatches 1217;  Indels    6;  Gaps    2;

Qy      2490 CAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGG 2549
             || | |  || |   ||   |  || |     ||||   |||| || ||| ||||| | |
Db      2845 CACCGGTCAGCCTGTCTTAAGTGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCG 2904

Qy      2550 TCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAAC 2609
             | || |||||||| ||||||||| | || || |||||||| |  ||   ||| ||  |
Db      2905 TGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATA 2964

Qy      2610 TGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTC 2669
             ||       || |  | ||   ||   | ||  |||  || |||||||| ||||| ||||
Db      2965 TGCGTATTTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTC 3024

Qy      2670 ATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTT 2729
             | |||   ||| | |  |     |||| ||| ||| |    |||||| |       |  |
Db      3025 ACCACGATGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACT 3084

Qy      2730 GGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTT 2789
             || | |  |||||||| || |||||  | |  || || || ||| || | |  || | ||
Db      3085 GGCCCGACTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTT 3144

Qy      2790 CCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGA 2849
               ||  | ||      ||| ||      | |  |      | | ||     |  ||||||
Db      3145 GTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGA 3204

Qy      2850 CAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAAC 2909
                | || || ||   | |||| ||  |       ||   | ||    | || |  |
Db      3205 TGGATACCGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTA 3264

Qy      2910 TTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGG 2969
             | | ||| |  | ||||||||| || || ||      | | |   ||   |||    |||
Db      3265 TCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGG 3324

Qy      2970 GTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGC 3029
             | |  |||| |||| |||| || || ||||   ||   || ||||| ||  |   |   |
Db      3325 GCTTCCAGGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTC 3384

Qy      3030 CGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGC 3089
             | ||  | ||||||||||| ||||||||||||   |||||||| || || ||||  ||||
Db      3385 CCATTATCATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGC 3444

Qy      3090 CGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGG 3149
             |||||    |     |  | || |||||||| || || ||  || |  |  |  |  | |
Db      3445 CGCTCTCGCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAG 3504

Qy      3150 CGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGA 3209
                   |   ||| || ||  | ||    ||||||||||   || || || |||  |||
Db      3505 ATCCTGAATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTC 3564

Qy      3210 GTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATC 3269
               || |||||||| ||||| || |||   ||||| ||||| |||   |||||   |||
Db      3565 AGACACTGCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACG 3624

Qy      3270 GATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATG 3329
              |||||||   |||||||||  ||  || || || ||  | || ||||| ||||| |||
Db      3625 CATTCTGG---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATC 3681

Qy      3330 CCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGT 3389
             | |||||||| || ||  |||||   || || ||  | || ||     |    || ||
Db      3682 CTACCACCGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTG 3741

Qy      3390 TGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGC 3449
                 |       |   | |  |   |||    ||    |  ||||||||  |||  ||||
Db      3742 ATGTCTTTACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGC 3801

Qy      3450 GCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGC 3509
                | ||| | ||||||||  |||| || |||||||||||||| |||||     |  |
Db      3802 AGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATT 3861

Qy      3510 AGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTT 3569
               |||||||| || ||||| ||||||||  |  |   ||  |  ||||||  ||| ||||
Db      3862 CACGCTGGCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTT 3921

Qy      3570 ACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGC 3629
             | || ||  ||||    |||  ||||      || |||||| ||||||||||||| || |
Db      3922 ATGCGGGAAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTC 3981

Qy      3630 CCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAA 3689
             |  |    | |||  |||  ||  |  |||   ||        | |   ||  | |||||
Db      3982 CGGTATTAGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAA 4041

Qy      3690 ACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGC 3749
             ||  |||   || ||||| |||||||| || |||||||| ||||| ||||| || |||||
Db      4042 ACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGC 4101

Qy      3750 GGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGT 3809
              ||||| || ||   |    | || || || || || ||||| ||||| || || |||||
Db      4102 TGCGGGGATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGT 4161

Qy      3810 GTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGC 3869
              |  |||||    || |||||||| ||||| || || ||||| |||||||||||||| ||
Db      4162 CTACGAGGATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGC 4221

Qy      3870 CATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCC 3929
             ||| |||||  ||||  | ||||||||||||  | | |  || || ||||| ||| ||||
Db      4222 CATTGAAGAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCC 4281

Qy      3930 GGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAA 3989
             ||| ||||| || || |||||||| |||||    || ||     || ||||||| ||  |
Db      4282 GGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGA 4341

Qy      3990 TCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCG 4049
             |||  |   ||| ||   ||||||      ||| ||||| |  ||||| ||||| |||||
Db      4342 TCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCG 4401

Qy      4050 TCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGT 4109
             ||| ||||| | ||||||||| |  || |||| ||| || ||||| || ||||| || ||
Db      4402 TCAGTTTCTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGT 4461

Qy      4110 GCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGC 4169
              || ||| |   | |  | |||||||| |||| ||| || ||  |||     ||||| ||
Db      4462 CCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGC 4521

Qy      4170 ATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGC 4229
             |||| | ||||| ||||| | ||| || | |  |  ||  | || || |||||||| |||||
Db      4522 ATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGC 4581

Qy      4230 GTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATAC 4289
             |||||| ||||||||  || |||||||||| ||||| || || || ||||| || |||||
Db      4582 GTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATAC 4641

Qy      4290 CTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGC 4349
              || || || |||||||||||||| |||||||| ||||||||||| ||||||||||||||
Db      4642 GTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGC 4701

Qy      4350 GCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGC 4409
               | || ||||| |||      |||||  ||||  |||| |||||  |  |  | || ||
Db      4702 ATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGC 4761

Qy      4410 CGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGG 4469
             |  ||| |||||  ||||||| |  || ||    |||   | ||| || ||||||| |||
Db      4762 CAGAGTCAGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGG 4821

Qy      4470 ACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAA 4529
              ||  | |||    |    || || || || ||||| || ||||| ||||||||   | |
Db      4822 TCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGA 4881

Qy      4530 ACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCAT 4589
              ||| |  |||| || ||    ||  | || |   | |     ||| ||  |||||||||
Db      4882 CCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCAT 4941

Qy      4590 GAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCA 4649
             ||| ||||| |||||||| |||||||||||  ||||||| |  ||||| |||||||||||
Db      4942 GAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCA 5001

Qy      4650 CCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAG 4709
              ||||| || ||||| ||||||||||| || ||| |||| || || ||||| ||
Db      5002 TCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGA 5061
```

```
Qy         4710 CCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCT 4769
                    || ||||| || || |||||||| ||||||||||||||   ||  ||| |||| ||
Db         5062 TGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACT 5121

Qy         4770 GGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGC 4829
                ||| ||    || |||||||| |  |||||||| ||||||||||||||| | ||||||||
Db         5122 GGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGC 5181

Qy         4830 TGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGA 4889
                  || ||||||||      || || ||     | ||||| || |||||||| || |  ||
Db         5182 GAACCTCGACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGA 5241

Qy         4890 GCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTT 4949
                |||||| |||||||| || |||||    ||  | || || ||| | ||||||||||||||
Db         5242 GCAGCATTACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTT 5301

Qy         4950 TGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTT 5009
                |||||||||||||  |||  ||||| |||| ||||| || |||||  |||||||  || |
Db         5302 TGATTTTCGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCT 5361

Qy         5010 GACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCT 5069
                 || || ||||||||||| || |||||||||||||||      |  |  |||||||| ||
Db         5362 TACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCT 5421

Qy         5070 GGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGC 5129
                ||| || || || |||||||| ||||| |||   || ||||| ||||| ||||| || ||
Db         5422 GGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGC 5481

Qy         5130 CACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAA 5189
                 || ||  |  |||||||||||||||        |||||   | | | |||||  | ||
Db         5482 GACAGCAGGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCAT 5538

Qy         5190 GTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGAT 5249
                |      |||| ||||| || |||||||| ||||| || || ||| | ||| | ||||||
Db         5539 GCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGAT 5598

Qy         5250 GCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTG 5309
                |||   || || ||||||   |||||||| || || ||||| || || || || |||||
Db         5599 GCAAAAACCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTT 5658

Qy         5310 ATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCA 5369
                ||||| |  || |||||||||      | |  ||||| |||| |  ||    |  |   |
Db         5659 ATTGACGATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAA 5718

Qy         5370 CGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAA 5429
                ||| ||  ||| ||  | ||  |||| ||||  |||||    ||  |||| ||| | |||
Db         5719 CGTCTGATGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAA 5778

Qy         5430 CCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCG 5489
                |||||||||  ||| |||||||||||||| ||   |||  |  |  | |        |||
Db         5779 CCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCG 5838

Qy         5490 TTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGAT 5549
                ||||||||||||||||||||||| ||||||| |||| |||  || |||       | |||
Db         5839 TTTGATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGAT 5898

Qy         5550 GACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTA 5609
                || |||||||| || ||||||||||| || ||||| ||||| |||||||||||||   |
Db         5899 GATACGCTGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATC 5958

Qy         5610 ATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAG 5669
                ||||||||||| ||  | ||   ||| |  | |||||| ||  | || ||||| || |||
Db         5959 ATGGGCGTGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAG 6018

Qy         5670 CTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCG 5729
                 | || || ||||| ||||| ||||| || ||||| |    || ||||| |||||||||
Db         6019 TTGACCAATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCG 6078

Qy         5730 CAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGT 5789
                  |  |||||| || || |  ||  |   |   || |      | |||||  || | |||
Db         6079 GCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGT 6138

Qy         5790 GCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCG 5849
                   |||||   |    | ||    ||||| |||  ||| || ||||| || |||   ||
Db         6139 CAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCT 6198

Qy         5850 GCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGC 5909
                ||   | ||||  |  | || || || ||||| || || ||||| |||||||||||
Db         6199 GCCACAGCCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAG 6258

Qy         5910 GGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCA 5969
                   ||||| ||  |||| || ||||||||| | || ||||||  ||| |||||||||
Db         6259 CAGGTTTACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAG 6318

Qy         5970 CGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCG 6029
                ||||||   || |  |     |||||| |  |||||||| ||   |   | |||  |||
Db         6319 CGGCAGTCAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCC 6378

Qy         6030 ATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTC 6087
                 | ||||| ||  ||     || |   || || || ||   ||| |||||||| || |
Db         6379 CTTACTTCCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC 6436

RESULT 10
ACF67367_41/c
Continuation (42 of 57) of ACF67367 from base 4100001 (Photorhabdus luminescens nucleot
WP Sequence split into 57 fragments  LOCUS ACF67367 Accession Acf67367
WP      Fragment Name        Begin       End
WP      ACF67367_00              1    110000
WP      ACF67367_01         100001    210000
WP      ACF67367_02         200001    310000
WP      ACF67367_03         300001    410000
WP      ACF67367_04         400001    510000
WP      ACF67367_05         500001    610000
WP      ACF67367_06         600001    710000
WP      ACF67367_07         700001    810000
WP      ACF67367_08         800001    910000
WP      ACF67367_09         900001   1010000
WP      ACF67367_10        1000001   1110000
WP      ACF67367_11        1100001   1210000
WP      ACF67367_12        1200001   1310000
WP      ACF67367_13        1300001   1410000
WP      ACF67367_14        1400001   1510000
WP      ACF67367_15        1500001   1610000
WP      ACF67367_16        1600001   1710000
WP      ACF67367_17        1700001   1810000
WP      ACF67367_18        1800001   1910000
WP      ACF67367_19        1900001   2010000
WP      ACF67367_20        2000001   2110000
WP      ACF67367_21        2100001   2210000
WP      ACF67367_22        2200001   2310000
WP      ACF67367_23        2300001   2410000
WP      ACF67367_24        2400001   2510000
WP      ACF67367_25        2500001   2610000
WP      ACF67367_26        2600001   2710000
WP      ACF67367_27        2700001   2810000
WP      ACF67367_28        2800001   2910000
WP      ACF67367_29        2900001   3010000
WP      ACF67367_30        3000001   3110000
WP      ACF67367_31        3100001   3210000
WP      ACF67367_32        3200001   3310000
WP      ACF67367_33        3300001   3410000
WP      ACF67367_34        3400001   3510000
WP      ACF67367_35        3500001   3610000
WP      ACF67367_36        3600001   3710000
WP      ACF67367_37        3700001   3810000
WP      ACF67367_38        3800001   3910000
WP      ACF67367_39        3900001   4010000
WP      ACF67367_40        4000001   4110000
WP      ACF67367_41        4100001   4210000
WP      ACF67367_42        4200001   4310000
WP      ACF67367_43        4300001   4410000
WP      ACF67367_44        4400001   4510000
WP      ACF67367_45        4500001   4610000
WP      ACF67367_46        4600001   4710000
WP      ACF67367_47        4700001   4810000
WP      ACF67367_48         4800001     4910000
WP      ACF67367_49         4900001     5010000
WP      ACF67367_50         5000001     5110000
WP      ACF67367_51         5100001     5210000
WP      ACF67367_52         5200001     5310000
WP      ACF67367_53         5300001     5410000
WP      ACF67367_54         5400001     5510000
WP      ACF67367_55         5500001     5610000
WP      ACF67367_56         5600001     5648894

  Query Match              19.6%;  Score 1374.8;  DB 10;  Length 110000;
  Best Local Similarity   62.0%;  Pred. No. 0;
  Matches 2211;  Conservative    0;  Mismatches 1347;  Indels    8;  Gaps    2;

Qy       2532 AATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGC 2591
              | |  ||| ||||| |||| || |  ||| |||| || ||  |||| ||  ||||| | |
Db      23257 ACTGGGATCTGATTCTGGTGGGAGGAGGATTGGCCAACGGACTGATCGCAATGCGTTTCC 23198

Qy       2592 GTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCA 2651
                || |   | ||||| |||  |    || |  | || | |   |        | |||||
Db      23197 AGCAGTGTAAACCACATCTGCGAGTGTTGCTTATTGAAAACACAGAAACAATAGGAGGCA 23138

Qy       2652 ATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAAC 2711
              |||| || ||||| ||||||||    ||||| |  |  |   ||| |   ||| |   ||
Db      23137 ATCACACGTGGTCATTTCATCAACATGATCTTACTGAGGCGGAACATGAGTGGATAGCAC 23078

Qy       2712 CGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATC 2771
              |||||||||||   || |||||||||||  | ||   |||||| ||  | |  ||||
Db      23077 CGCTGATTACCTATCGCTGGTCAGGTTACGACGTCATTTTTCCAGCATTTCAACGCACAT 23018

Qy       2772 TGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGA 2831
              ||         ||||     ||| ||||     | |||||  ||   ||  | ||
Db      23017 TGCCACATTCATATTTCAGTATCACATCCCAACACTTTGCAAGCATACTCCATGCATATT 22958

Qy       2832 TGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGC 2891
              |||| ||   | |   |||      |   |||||  || | |  ||   | | ||
Db      22957 TGGGCGAACGTATACAGACTCGTTTATTGGTACAGGAGCTGACTCCACAGAAAGTTTACT 22898

Qy       2892 TGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGA 2951
              |   ||| |||      ||   |||    |  || ||||||||  | |||  ||  ||
Db      22897 TACAGGACGGCTCGTCTCTAAGTGCTGGCGCAGTCATTGATGGGAGAGGCTGGCGACCAG 22838

Qy       2952 CGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGC 3011
                |||  | |   |   ||    ||||  ||| |||| || || ||| ||||||   ||
Db      22837 GACCATTTATAGGGAGTGGCACCCAGGCATTTTTTGGTCAGGAATGGGAGCTGGAAGAGT 22778

Qy       3012 CGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATC 3071
              |||||    |    || |||||  | |||||| | |  || |  ||| |  | |||||||
Db      22777 CGCACTCTTTAACCCACCCGATTTTAATGGATACTAGTGTGGGACAGGATACAGGTTATC 22718

Qy       3072 GTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACG 3131
              | ||| ||||    ||||||| ||  | |   ||| ||  |||| || || || |||||||
Db      22717 GATTTATCTATGTCCTGCCGTTCTCCTCAACTCGTCTGCTGATAGAGGACACTCATTACG 22658

Qy       3132 TTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATC 3191
              || | | |    | |  |   ||       | |        ||||| || || || ||
Db      22657 TTGATCGGGGGCCACCTGATAAGGCTTTGTCGCAGGCTACTATCGCAGAGTACGCGAAGA 22598

Qy       3192 AGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCC 3251
              | || || ||||  ||| |||  ||  | || ||||||  |||     | || ||||| |
Db      22597 AACATGGATGGAAACTGGGTAAACTCATTCGAGAAGAGAGCGGTTGTCTTCCAATTACAC 22538

Qy       3252 TGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGC 3311
              | |  ||  |  |       ||||||  |||||     || ||  |   | | ||  | |
Db      22537 TTACGGGAGATTTTACCTCTTTCTGGGCACAGCTAGCAGGACAGCCCACCTGTGGGTTAC 22478

Qy       3312 GCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGG 3371
              | || |   |||||||  |||| || || |||||  ||||   |||  |   | | || |
Db      22477 GTGCAGCTTTGTTTCACCCCACAACAGGCTACTCTCTGCCACACGCTATTCGGTTGGCAG 22418

Qy       3372 AGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCT 3431
              |     | |  || |||  |         |    | |  |   |       ||      |
Db      22417 ATCGTATTGTTGCTCTGCCGGAGCTTACCGATACCTCCTTATTCATTACCCTCAGGGATT 22358

Qy       3432 TTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTT 3491
                || || || |||||||   | ||||| ||||||||||| ||||| |||||||| ||
Db      22357 ACGCACGACAACAGTGGCAACACCAGCGCTTTTTCCGTCTTCTAAATCGCATGCTCTTCC 22298

Qy       3492 TGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCG 3551
              | || ||  | || || || || ||||  || ||||||||||| || |  ||     |
Db      22297 TCGCTGGGGATCCACAACAACGTTGGCAGGTAATGCAACGTTTCTATCAACTTTCCCCAA 22238

Qy       3552 GGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGT 3611
                 | |||   || ||||| || | ||||||         ||| ||  | ||||||||
Db      22237 ATCTGATTGCACGTTTTTATGCGGAGCAACTTAATTCCGTCGACAAGGCCCGGATTCTCA 22178

Qy       3612 GCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAA-----TTCTGT 3666
                || || || || ||||| ||    |  || ||    ||  |||| ||         |
Db      22177 TAGGTAAACCACCAGTGCCGATAAAGGGTGCTCTAAAGGCAATGTTTAAACAACACAAGA 22118

Qy       3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
               |   |||||     |  ||||      |    || ||||| || || ||||| || |||
Db      22117 AGCTTCAGGGTTTTTATTATGATTAAAGCGCTGGTAATTGGTGCTGGTTTTGGTGGGCTG 22058

Qy       3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
              || ||||| ||  | || ||  | ||||| || || ||     |  | ||||| || ||
Db      22057 GCACTGGCAATAAGGCTCCAGTCTGCGGGGATTCCGACATGTATTTTGGAGCAACGGGAT 21998

Qy       3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
              ||||| || |||||||| ||||| | | ||||    || || ||||| ||||||| |||
Db      21997 AAACCAGGTGGACGCGCTTATGTTTATAAGGAACAGGGATTCACCTTTGATGCCGGCCCC 21938

Qy       3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
              || || ||||||| ||| |  | ||| |||||| ||||| || | ||| || |||    |
Db      21937 ACCGTAATCACCGCTCCTAATGTCATTGAAGAATTGTTTACCCAGGCTGGTAAACGTATG 21878

Qy       3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
                 |||||||||||  |  | || |||   || || ||||| || || ||||    ||
Db      21877 GCTGATTACGTCGATTTACTTCCCGTACACCCTTTTTATCGGCTCTGTTGGGAGTCGGGT 21818

Qy       3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
              ||   | ||||||||||||| ||||||  |  |||||    || |||   |||||||||
Db      21817 AAGATGTTTGATTACGACAATGATCAGCAACATCTGGAAGCACAAATCCATACGTTCAAT 21758

Qy       4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
              ||   ||||||  | || ||||| |  |||||| | ||||| || ||||  ||||  ||
Db      21757 CCAAGAGATGTTAACGGGTATCGACGCTTTCTAGACTATTCCCGGGAAGCCTTTAATGAA 21698

Qy       4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
              || || |||||||| || |||||||| || ||     | || |||||| || | |  ||
Db      21697 GGCTACCTGAAACTTGGAACGGTGCCCTTCCTCTCTTTTCGCGACATGTTGAGTGCGGCT 21638

Qy       4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
              || ||||| | |||||| || |||||||| || || || ||   ||||||    |||||
Db      21637 CCACAGTTGATACGTCTACACGCATGGCGTAGTGTTTATAGTCAGGTGGCACGCTTTATC 21578

Qy       4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
               |||| ||    ||||| ||||| ||||| || |||||  |  || | || ||||| ||
Db      21577 AAGGATGAGAGCCTGCGCCAGGCATTTTCATTTCACTCGCTATTGATAGGAGGTAACCCC 21518

Qy       4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
              |||||  | || || |||||||| || || || || ||||| || |||||||| || |||
Db      21517 TTTGCTGCATCTTCTATCTATACTTTGATCCACGCACTGGAACGAGAATGGGGAGTCTGG 21458

Qy       4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
              |||||||| || || ||| |||| |||||  |||  |||| |  | ||||    ||| |
Db      21457 TTTCCGCGAGGAGGAACCAGCGCTCTGGTTGAGGCAATGGTGAAATTGTTTACTGACATT 21398

Qy       4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
              || || ||  |    || ||||||  | || |  |   |    |||   ||||||||  |
Db      21397 GGTGGAGAAATTGAGCTTAATGCAAAGGTAAAACGCTTCACCACCCATGGCAATCGGGTC 21338

Qy       4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
              |  || || || ||||  |  ||||||    |       | || ||||||||| ||||||
Db      21337 ACAGGTGTCCAATTAGCTGATGGACGAAATATGACATGTGACGTTGTGGCCTCAAATGCC 21278
```

```
Qy          4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
                 || ||  | ||||||||  |  | ||| |  | || |||||| | | |||    || ||
Db         21277 GATGTTATTCATACCTATAAGCACCTGTTAGGTCAGCATCCGGTCGGAATAGCCCGCGCT 21218

Qy          4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
                   | | |||  |||||| || |||||||||||| ||||| || | || ||||| || |||||
Db         21217 CGAACACTGGTGCGTAAACGAATGAGCAACTCACTGTTCGTGCTTTATTTCGGTCTGAAC 21158

Qy          4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
                 || |  |||   ||  | || || || || || || |||| || || |||   ||  |
Db         21157 CATCACCATACGCAATTAGCCCATCATACAGTTTGCTTTGGTCCACGCTATAAGGAACTC 21098

Qy          4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
                 ||||| || |||||  |      ||  ||| |||| ||||||||||| ||| || || ||
Db         21097 ATCGAAGATATTTTTTATCATGACCGACTGTCAGAGGATTTTTCACTCTATCTCCATGCC 21038

Qy          4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
                 ||||     |  || || || || || || |  || ||||  || |  || ||  | || 
Db         21037 CCCTCTGTTACTGACCCATCCCTTGCTCCCCGAGGATGCGCAAGTTACTACGTTCTGGCC 20978

Qy          4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
                 || || || ||| | || ||| |  |  ||||| |    || || || ||||| ||
Db         20977 CCTGTACCACATTTGGGGACAGCTAATTTAAACTGGGATATAGAGGGGCCACGCTTACGT 20918

Qy          4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
                 || || ||||| || |||||||||  || |||| ||||| ||  | |   |||| |||||
Db         20917 GACCGTATTTTCGCGTATCTGGAAAAGTACTATATGCCCGGTCTGCTCAAGCAGTTAGTC 20858

Qy          4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
                    ||  | || ||||| || ||||||||||| || |||   ||  |||||||||| ||| |||
Db         20857 GTTCATCGTATTTTTACCCCATTTGATTTTCGTGACCAACTCAATGCCCATCTCGGTTCG 20798

Qy          4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
                 || || ||  | || ||| |  ||||  |||||| |||||||| || || || || |
Db         20797 GCCTTCTCCTTTGAACCGTTGCTGACACAAAGCGCTTGGTTCCGACCACACAATCGGGAC 20738

Qy          5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
                   |   ||    ||||| |||||| || || || || ||||||| || || ||  | |||
Db         20737 AACCGCATTGATAATCTGTATCTCGTCGGCGCAGGCACGCATCCTGGAGCAGGTATTCCC 20678

Qy          5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
                 |||||||| || || || || ||||||||||   ||||| ||||||  |||  |||||
Db         20677 GGCGTGATTGGATCAGCAAAAGCCACCGCCACATTGATGTTAGAGGATATCGCTAAATGA 20618

Qy          5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
                 |||    ||   ||||| | || || || || | |  |||| |   |||||  || ||||| |
Db         20617 ATCCTGCGC---TACTTAAACACGTTACTCAGATAATGGAGCAAGGCTCAAAAAGTTTTG 20561

Qy          5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
                 |||  | | ||   || || ||| || | || || | ||||||  |||||||||||| || |
Db         20560 CCAGTGTCACCCGACTTTTCGATACAGCAACACGACACAGCACAATGATGCTGTACGCCT 20501

Qy          5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
                 |||| ||| | || |||||| | || || |||||  ||||||   |      |
Db         20500 GGTGCCGTTATTGTGATGATATAATAGACGGGCAAGAGCTGGGGAGACAAATAAGCAGTG 20441

Qy          5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
                  ||      |   ||| |  |      | ||    |||||      ||    |  || |
Db         20440 TTGATAAGTATAGCGCCCGCGAGAAACTTCAAATGCTGCAATACCTGACAAAGCAGGCTT 20381

Qy          5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
                 ||  |||   ||  |||   ||||| || |||   || |||||  ||| | | |
Db         20380 ACGACGGTTTGCCGATGACCGAACCTGCTTTTGCAGCCTTTCAGACAGTAGCTCTGAGTA 20321

Qy          5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
                 |  |  | || || |||| ||| || || || ||||||||||| ||||| |||||||| || |
Db         20320 ATGAAATCCCTCAGCAACAGGCCTTCGAACATCTGGAAGGATTCGCAATGGATGTACTCT 20261

Qy          5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
                   || | ||||   |  || ||||| || || | || || ||| || ||||||| || |
Db         20260 GTGAGCCTTACAGAACGTTGGATGATACTCTGAGATATTGCTACCATGTCGCGGGAGTAG 20201

Qy          5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
                 | ||  ||||||||||  | || ||||| || ||||| |     ||||| ||||  ||||
Db         20200 TTGGGCTGATGATGGCAAGGGTTATGGGGGTTCGCGAGGCGAGTGTGCTGGATCGGGCCT 20141

Qy          5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
                 ||||| ||||| | || || ||| | |||||||||||| || || ||| | || |||||
Db         20140 GCGATCTAGGAATTGCTTTTCAGTTAACTAACATTGCTCGGGATATTATTGAGGATGCGA 20081

Qy          5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
                 |    || || |||||||| |||| |   |||||  | || |  ||  ||    | || |
Db         20080 AGGCCGGGCGTTGCTATCTTCCGCTAGAATGGCTTCACCAAGAAGGGCTAATGCCAGACA 20021

Qy          5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                 | || | |    |  || | ||| | ||||| ||     ||||| | || ||  | | ||
Db         20020 CACTTATTTATACCGAAAACCGTCCTGCGCTAGCTCGGGTGGCATCCCGATTGATTGTGG 19961

Qy          5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
                 | |||||| | ||||||  ||| || |   | ||  |    ||||||||  |||| || |
Db         19960 AAGCGGAATCTTATTATACCTCCGCTCTGACAGGGCTTGTTGGTTTACCTTTGCGTTCAG 19901

Qy          5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
                 | |||| |||||||| | |||  ||  | ||||| ||||| ||| | ||||| || || |
Db         19900 CCTGGGTCATCGCTTCTGCTCATGGTATCTATCGTGAAATCGGCATTAAAGTGCAACAAG 19841

Qy          5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
                 | || || |  |||||||| |   | ||  | || | |    || || ||   |  |  |
Db         19840 CAGGAGTACGGGCCTGGGAATGCAGACAAAGAACTAATCGTGGAGAGAAGGTTACCTTAT 19781

Qy          6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
                 |  ||   || |||     |||| | |  || || ||| |     |    || ||
Db         19780 TAATGGCTGGCGCAATGAAGGCGCTTATATCTCGGGTGGCGATAACAACGCCTCGCGATC 19721

Qy          6067 CTGGTCTGTGGCAGCGTCCTCGTTGA 6092
                 |     |||||||||| |||| || |
Db         19720 CCAAATTGTGGCAGCGACCTCATTAA 19695


RESULT 11
ACF65388_06
Continuation (7 of 13) of ACF65388 from base 600001 (Photorhabdus luminescens nucleotid
WP Sequence split into 13 fragments  LOCUS ACF65388 Accession Acf65388
WP     Fragment Name       Begin      End
WP     ACF65388_00             1    110000
WP     ACF65388_01        100001    210000
WP     ACF65388_02        200001    310000
WP     ACF65388_03        300001    410000
WP     ACF65388_04        400001    510000
WP     ACF65388_05        500001    610000
WP     ACF65388_06        600001    710000
WP     ACF65388_07        700001    810000
WP     ACF65388_08        800001    910000
WP     ACF65388_09        900001   1010000
WP     ACF65388_10       1000001   1110000
WP     ACF65388_11       1100001   1210000
WP     ACF65388_12       1200001   1225559

  Query Match             19.6%;  Score 1374.8;  DB 10;  Length 110000;
  Best Local Similarity   62.0%;  Pred. No. 0;
  Matches 2211;  Conservative    0;  Mismatches 1347;  Indels    8;  Gaps    2;

Qy          2532 AATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGC 2591
                 | |  ||| ||||| |||| || |  ||| |||| || ||  |||| ||  ||||| | |
Db        101680 ACTGGGATCTGATTCTGGTGGGAGGAGGATTGGCCAACGGACTGATCGCAATGCGTTTCC 101739

Qy          2592 GTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCA 2651
                   || |   | ||||| |||  |    || |  | || | |   |        |  ||||
Db        101740 AGCAGTGTAAACCACATCTGCGAGTGTTGCTTATTGAAAACACAGAAACAATAGGAGGCA 101799

Qy          2652 ATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAAC 2711
                 |||| || ||||| ||||||||    ||||| |  |  |   ||| |   ||| |   ||
Db        101800 ATCACACGTGGTCATTTCATCAACATGATCTTACTGAGGCGGAACATGAGTGGATAGCAC 101859

Qy          2712 CGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATC 2771
                 |||||||||||   || |||||||||||  | ||   |||||| ||  | |  ||||
Db   101860 CGCTGATTACCTATCGCTGGTCAGGTTACGACGTCATTTTTCCAGCATTTCAACGCACAT 101919

Qy     2772 TGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGA 2831
Db   101920 TGCCACATTCATATTTCAGTATCACATCCCAACACTTTGCAAGCATACTCCATGCATATT 101979

Qy     2832 TGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGC 2891
Db   101980 TGGGCGAACGTATACAGACTCGTTTATTGGTACAGGAGCTGACTCCACAGAAAGTTTACT 102039

Qy     2892 TGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGA 2951
Db   102040 TACAGGACGGCTCGTCTCTAAGTGCTGGCGCAGTCATTGATGGGAGAGGCTGGCGACCAG 102099

Qy     2952 CGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGC 3011
Db   102100 GACCATTTATAGGGAGTGGCACCCAGGCATTTTTTGGTCAGGAATGGGAGCTGGAAGAGT 102159

Qy     3012 CGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATC 3071
Db   102160 CGCACTCTTTAACCCACCCGATTTTAATGGATACTAGTGTGGGACAGGATACAGGTTATC 102219

Qy     3072 GTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACG 3131
Db   102220 GATTTATCTATGTCCTGCCGTTCTCCTCAACTCGTCTGCTGATAGAGGACACTCATTACG 102279

Qy     3132 TTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATC 3191
Db   102280 TTGATCGGGGGCCACCTGATAAGGCTTTGTCGCAGGCTACTATCGCAGAGTACGCGAAGA 102339

Qy     3192 AGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCC 3251
Db   102340 AACATGGATGGAAACTGGGTAAACTCATTCGAGAAGAGAGCGGTTGTCTTCCAATTACAC 102399

Qy     3252 TGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGC 3311
Db   102400 TTACGGGAGATTTTACCTCTTTCTGGGCACAGCTAGCAGGACAGCCCACCTGTGGGTTAC 102459

Qy     3312 GCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGG 3371
Db   102460 GTGCAGCTTTGTTTCACCCCACAACAGGCTACTCTCTGCCACACGCTATTCGGTTGGCAG 102519

Qy     3372 AGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCT 3431
Db   102520 ATCGTATTGTTGCTCTGCCGGAGCTTACCGATACCTCCTTATTCATTACCCTCAGGGATT 102579

Qy     3432 TTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTT 3491
Db   102580 ACGCACGACAACAGTGGCAACACCAGCGCTTTTTCCGTCTTCTAAATCGCATGCTCTTCC 102639

Qy     3492 TGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCG 3551
Db   102640 TCGCTGGGGATCCACAACAACGTTGGCAGGTAATGCAACGTTTCTATCAACTTTCCCCAA 102699

Qy     3552 GGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGT 3611
Db   102700 ATCTGATTGCACGTTTTTATGCGGAGCAACTTAATTCCGTCGACAAGGCCCGGATTCTCA 102759

Qy     3612 GCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAA-----TTCTGT 3666
Db   102760 TAGGTAAACCACCAGTGCCGATAAAGGGTGCTCTAAAGGCAATGTTTAAACAACACAAGA 102819

Qy     3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
Db   102820 AGCTTCAGGGTTTTTATTATGATTAAAGCGCTGGTAATTGGTGCTGGTTTTGGTGGGCTG 102879

Qy     3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
Db   102880 GCACTGGCAATAAGGCTCCAGTCTGCGGGGATTCCGACATGTATTTTGGAGCAACGGGAT 102939

Qy     3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
Db   102940 AAACCAGGTGGACGCGCTTATGTTTATAAGGAACAGGGATTCACCTTTGATGCCGGCCCC 102999
```

```
Qy     3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
Db   103000 ACCGTAATCACCGCTCCTAATGTCATTGAAGAATTGTTTACCCAGGCTGGTAAACGTATG 103059

Qy     3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
Db   103060 GCTGATTACGTCGATTTACTTCCCGTACACCCTTTTTATCGGCTCTGTTGGGAGTCGGGT 103119

Qy     3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
Db   103120 AAGATGTTTGATTACGACAATGATCAGCAACATCTGGAAGCACAAATCCATACGTTCAAT 103179

Qy     4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
Db   103180 CCAAGAGATGTTAACGGGTATCGACGCTTTCTAGACTATTCCCGGGAAGCCTTTAATGAA 103239

Qy     4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
Db   103240 GGCTACCTGAAACTTGGAACGGTGCCCTTCCTCTCTTTTCGCGACATGTTGAGTGCGGCT 103299

Qy     4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
Db   103300 CCACAGTTGATACGTCTACACGCATGGCGTAGTGTTTATAGTCAGGTGGCACGCTTTATC 103359

Qy     4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
Db   103360 AAGGATGAGAGCCTGCGCCAGGCATTTTCATTTCACTCGCTATTGATAGGAGGTAACCCC 103419

Qy     4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
Db   103420 TTTGCTGCATCTTCTATCTATACTTTGATCCACGCACTGGAACGAGAATGGGGAGTCTGG 103479

Qy     4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
Db   103480 TTTCCGCGAGGAGGAACCAGCGCTCTGGTTGAGGCAATGGTGAAATTGTTTACTGACATT 103539

Qy     4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
Db   103540 GGTGGAGAAATTGAGCTTAATGCAAAGGTAAAACGCTTCACCACCCATGGCAATCGGGTC 103599

Qy     4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
Db   103600 ACAGGTGTCCAATTAGCTGATGGACGAAATATGACATGTGACGTTGTGGCCTCAAATGCC 103659

Qy     4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
Db   103660 GATGTTATTCATACCTATAAGCACCTGTTAGGTCAGCATCCGGTCGGAATAGCCCGCGCT 103719

Qy     4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
Db   103720 CGAACACTGGTGCGTAAACGAATGAGCAACTCACTGTTCGTGCTTTATTTCGGTCTGAAC 103779

Qy     4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
Db   103780 CATCACCATACGCAATTAGCCCATCATACAGTTTGCTTTGGTCCACGCTATAAGGAACTC 103839

Qy     4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
Db   103840 ATCGAAGATATTTTTTATCATGACCGACTGTCAGAGGATTTTTCACTCTATCTCCATGCC 103899

Qy     4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
Db   103900 CCCTCTGTTACTGACCCATCCCTTGCTCCCCGAGGATGCGCAAGTTACTACGTTCTGGCC 103959

Qy     4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
Db   103960 CCTGTACCACATTTGGGGACAGCTAATTTAAACTGGGATATAGAGGGGCCACGCTTACGT 104019

Qy     4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
Db   104020 GACCGTATTTTCGCGTATCTGGAAAAGTACTATATGCCCGGTCTGCTCAAGCAGTTAGTC 104079
```

```
Qy        4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
                  ||  | || ||||| || ||||||||||| |||   ||  ||||||||| ||| |||
Db      104080 GTTCATCGTATTTTTACCCCATTTGATTTTCGTGACCAACTCAATGCCCATCTCGGTTCG 104139

Qy        4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
               || || ||  | || ||| |  |||| |||||||| |||||||| || || || || ||
Db      104140 GCCTTCTCCTTTGAACCGTTGCTGACACAAAGCGCTTGGTTCCGACCACACAATCGGGAC 104199

Qy        5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
                 |   ||    ||||| ||||| || || || || |||||||| || || ||  | |||
Db      104200 AACCGCATTGATAATCTGTATCTCGTCGGCGCAGGCACGCATCCTGGAGCAGGTATTCCC 104259

Qy        5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
               |||||||| || || || || ||||||||||   ||||| | ||||||  |||  |||||
Db      104260 GGCGTGATTGGATCAGCAAAAGCCACCGCCACATTGATGTTAGAGGATATCGCTAAATGA 104319

Qy        5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
               |||    ||   ||||| | || || || || |  |||| |   ||||| || ||||| |
Db      104320 ATCCTGCGC---TACTTAAACACGTTACTCAGATAATGGAGCAAGGCTCAAAAAGTTTTG 104376

Qy        5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
               |||  | | ||   || || ||| || | || || | ||||||  |||||||||| || |
Db      104377 CCAGTGTCACCCGACTTTTCGATACAGCAACACGACACAGCACAATGATGCTGTACGCCT 104436

Qy        5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
               |||| ||| | || |||||| | || || ||||||  ||||||   |      |
Db      104437 GGTGCCGTTATTGTGATGATATAATAGACGGGCAAGAGCTGGGGAGACAAATAAGCAGTG 104496

Qy        5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
                ||      |   ||| |  |      | ||    ||||||      ||    |  || |
Db      104497 TTGATAAGTATAGCGCCCGCGAGAAACTTCAAATGCTGCAATACCTGACAAAGCAGGCTT 104556

Qy        5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
               ||  |||   ||  |||   ||||| || |||   || ||||||  ||| ||  | | |
Db      104557 ACGACGGTTTGCCGATGACCGAACCTGCTTTTGCAGCCTTTCAGACAGTAGCTCTGAGTA 104616

Qy        5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
               |  |  | || || |||| ||| || || ||||||||||| ||||| ||||||| || |
Db      104617 ATGAAATCCCTCAGCAACAGGCCTTCGAACATCTGGAAGGATTCGCAATGGATGTACTCT 104676

Qy        5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
                 || | ||||   |  || ||||| || ||| | || ||||| || |||||||| || |
Db      104677 GTGAGCCTTACAGAACGTTGGATGATACTCTGAGATATTGCTACCATGTCGCGGGAGTAG 104736

Qy        5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
               | ||  ||||||||||  | || ||||| || ||||| |     ||||| ||||  ||||
Db      104737 TTGGGCTGATGATGGCAAGGGTTATGGGGGTTCGCGAGGCGAGTGTGCTGGATCGGGCCT 104796

Qy        5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
               ||||| ||||| | || || ||| | ||||||||||| || || ||| | || |||||
Db      104797 GCGATCTAGGAATTGCTTTTCAGTTAACTAACATTGCTCGGGATATTATTGAGGATGCGA 104856

Qy        5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
               |    || || |||||||| |||| |   |||||  | || |  ||  ||    | || |
Db      104857 AGGCCGGGCGTTGCTATCTTCCGCTAGAATGGCTTCACCAAGAAGGGCTAATGCCAGACA 104916

Qy        5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
               | || | |    |  || | ||| | ||||| ||     ||||| | || ||  | | ||
Db      104917 CACTTATTTATACCGAAAACCGTCCTGCGCTAGCTCGGGTGGCATCCCGATTGATTGTGG 104976

Qy        5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
               | |||||| | ||||||  ||| || |   | ||  |    ||||||||  |||| || |
Db      104977 AAGCGGAATCTTATTATACCTCCGCTCTGACAGGGCTTGTTGGTTTACCTTTGCGTTCAG 105036

Qy        5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
               | |||| ||||||| | ||||  ||  | ||||| ||||| ||| | ||||| || || |
Db      105037 CCTGGGTCATCGCTTCTGCTCATGGTATCTATCGTGAAATCGGCATTAAAGTGCAACAAG 105096

Qy        5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
               | || || |  |||||||| |   | ||  | || | |   || || ||  |  |  |
Db      105097 CAGGAGTACGGGCCTGGGAATGCAGACAAAGAACTAATCGTGGAGAGAAGGTTACCTTAT 105156

Qy        6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
```

```
               |  ||   || |||     |||| | |  || || ||| |     |    || ||
Db      105157 TAATGGCTGGCGCAATGAAGGCGCTTATATCTCGGGTGGCGATAACAACGCCTCGCGATC 105216

Qy        6067 CTGGTCTGTGGCAGCGTCCTCGTTGA 6092
               |     |||||||||| |||| || |
Db      105217 CCAAATTGTGGCAGCGACCTCATTAA 105242


RESULT 12
ACF65388_07
Continuation (8 of 13) of ACF65388 from base 700001 (Photorhabdus luminescens nucleotide
WP Sequence split into 13 fragments  LOCUS ACF65388 Accession Acf65388
WP      Fragment Name       Begin      End
WP      ACF65388_00             1    110000
WP      ACF65388_01        100001    210000
WP      ACF65388_02        200001    310000
WP      ACF65388_03        300001    410000
WP      ACF65388_04        400001    510000
WP      ACF65388_05        500001    610000
WP      ACF65388_06        600001    710000
WP      ACF65388_07        700001    810000
WP      ACF65388_08        800001    910000
WP      ACF65388_09        900001   1010000
WP      ACF65388_10       1000001   1110000
WP      ACF65388_11       1100001   1210000
WP      ACF65388_12       1200001   1225559

  Query Match             19.6%;  Score 1374.8;  DB 10;  Length 110000;
  Best Local Similarity   62.0%;  Pred. No. 0;
  Matches 2211;  Conservative    0;  Mismatches 1347;  Indels    8;  Gaps    2;

Qy        2532 AATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGC 2591
               | |  ||| ||||| |||| || |  ||| |||| || ||  |||| ||  ||||| | |
Db        1680 ACTGGGATCTGATTCTGGTGGGAGGAGGATTGGCCAACGGACTGATCGCAATGCGTTTCC 1739

Qy        2592 GTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCA 2651
                 || |   | ||||| |||  |    || |  | || | |   |        | |||||
Db        1740 AGCAGTGTAAACCACATCTGCGAGTGTTGCTTATTGAAAACACAGAAACAATAGGAGGCA 1799

Qy        2652 ATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAAC 2711
               |||| || ||||| ||||||||    ||||| |  |  |   ||| |   ||| |   ||
Db        1800 ATCACACGTGGTCATTTCATCAACATGATCTTACTGAGGCGGAACATGAGTGGATAGCAC 1859

Qy        2712 CGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATC 2771
               |||||||||||   || |||||||||||  | ||   |||||| ||  | |  ||||
Db        1860 CGCTGATTACCTATCGCTGGTCAGGTTACGACGTCATTTTTCCAGCATTTCAACGCACAT 1919

Qy        2772 TGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGA 2831
               ||         ||||     ||| ||||     | |||||  ||   ||  | ||
Db        1920 TGCCACATTCATATTTCAGTATCACATCCCAACACTTTGCAAGCATACTCCATGCATATT 1979

Qy        2832 TGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGC 2891
               |||| ||   | |   |||      |   |||||  || | |  ||   | | ||
Db        1980 TGGGCGAACGTATACAGACTCGTTTATTGGTACAGGAGCTGACTCCACAGAAAGTTTACT 2039

Qy        2892 TGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGA 2951
               |   ||| |||      ||   |||    |  || ||||||||  | |||  ||  ||
Db        2040 TACAGGACGGCTCGTCTCTAAGTGCTGGCGCAGTCATTGATGGGAGAGGCTGGCGACCAG 2099

Qy        2952 CGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGC 3011
                 |||  | |   |   ||    ||||  ||| |||| || || ||| ||||||   ||
Db        2100 GACCATTTATAGGGAGTGGCACCCAGGCATTTTTTGGTCAGGAATGGGAGCTGGAAGAGT 2159

Qy        3012 CGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATC 3071
               |||||    |    || |||||  | |||||| | |  || |  ||| |  | |||||||
Db        2160 CGCACTCTTTAACCCACCCGATTTTAATGGATACTAGTGTGGGACAGGATACAGGTTATC 2219

Qy        3072 GTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACG 3131
               | ||| ||||    |||||| ||  | |   ||| ||  |||| || || || |||||||
Db        2220 GATTTATCTATGTCCTGCCGTTCTCCTCAACTCGTCTGCTGATAGAGGACACTCATTACG 2279

Qy        3132 TTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATC 3191
               || | | |    | |  |   ||       | |        ||||| || || || ||
Db       2280 TTGATCGGGGGCCACCTGATAAGGCTTTGTCGCAGGCTACTATCGCAGAGTACGCGAAGA 2339

Qy       3192 AGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCC 3251

Db       2340 AACATGGATGGAAACTGGGTAAACTCATTCGAGAAGAGAGCGGTTGTCTTCCAATTACAC 2399

Qy       3252 TGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGC 3311

Db       2400 TTACGGGAGATTTTACCTCTTTCTGGGCACAGCTAGCAGGACAGCCCACCTGTGGGTTAC 2459

Qy       3312 GCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGG 3371

Db       2460 GTGCAGCTTTGTTTCACCCCACAACAGGCTACTCTCTGCCACACGCTATTCGGTTGGCAG 2519

Qy       3372 AGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCT 3431

Db       2520 ATCGTATTGTTGCTCTGCCGGAGCTTACCGATACCTCCTTATTCATTACCCTCAGGGATT 2579

Qy       3432 TTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTT 3491

Db       2580 ACGCACGACAACAGTGGCAACACCAGCGCTTTTTCCGTCTTCTAAATCGCATGCTCTTCC 2639

Qy       3492 TGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCG 3551

Db       2640 TCGCTGGGGATCCACAACAACGTTGGCAGGTAATGCAACGTTTCTATCAACTTTCCCCAA 2699

Qy       3552 GGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGT 3611

Db       2700 ATCTGATTGCACGTTTTTATGCGGAGCAACTTAATTCCGTCGACAAGGCCCGGATTCTCA 2759

Qy       3612 GCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAA-----TTCTGT 3666

Db       2760 TAGGTAAACCACCAGTGCCGATAAAGGGTGCTCTAAAGGCAATGTTTAAACAACACAAGA 2819

Qy       3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726

Db       2820 AGCTTCAGGGTTTTTATTATGATTAAAGCGCTGGTAATTGGTGCTGGTTTTGGTGGGCTG 2879

Qy       3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786

Db       2880 GCACTGGCAATAAGGCTCCAGTCTGCGGGGATTCCGACATGTATTTTGGAGCAACGGGAT 2939

Qy       3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846

Db       2940 AAACCAGGTGGACGCGCTTATGTTTATAAGGAACAGGGATTCACCTTTGATGCCGGCCCC 2999

Qy       3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906

Db       3000 ACCGTAATCACCGCTCCTAATGTCATTGAAGAATTGTTTACCCAGGCTGGTAAACGTATG 3059

Qy       3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966

Db       3060 GCTGATTACGTCGATTTACTTCCCGTACACCCTTTTTATCGGCTCTGTTGGGAGTCGGGT 3119

Qy       3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026

Db       3120 AAGATGTTTGATTACGACAATGATCAGCAACATCTGGAAGCACAAATCCATACGTTCAAT 3179

Qy       4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086

Db       3180 CCAAGAGATGTTAACGGGTATCGACGCTTTCTAGACTATTCCCGGGAAGCCTTTAATGAA 3239

Qy       4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146

Db       3240 GGCTACCTGAAACTTGGAACGGTGCCCTTCCTCTCTTTTCGCGACATGTTGAGTGCGGCT 3299

Qy       4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206

Db       3300 CCACAGTTGATACGTCTACACGCATGGCGTAGTGTTTATAGTCAGGTGGCACGCTTTATC 3359

Qy       4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266

Db       3360 AAGGATGAGAGCCTGCGCCAGGCATTTTCATTTCACTCGCTATTGATAGGAGGTAACCCC 3419
```

```
Qy       4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326

Db       3420 TTTGCTGCATCTTCTATCTATACTTTGATCCACGCACTGGAACGAGAATGGGGAGTCTGG 3479

Qy       4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386

Db       3480 TTTCCGCGAGGAGGAACCAGCGCTCTGGTTGAGGCAATGGTGAAATTGTTTACTGACATT 3539

Qy       4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446

Db       3540 GGTGGAGAAATTGAGCTTAATGCAAAGGTAAAACGCTTCACCACCCATGGCAATCGGGTC 3599

Qy       4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506

Db       3600 ACAGGTGTCCAATTAGCTGATGGACGAAATATGACATGTGACGTTGTGGCCTCAAATGCC 3659

Qy       4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566

Db       3660 GATGTTATTCATACCTATAAGCACCTGTTAGGTCAGCATCCGGTCGGAATAGCCCGCGCT 3719

Qy       4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626

Db       3720 CGAACACTGGTGCGTAAACGAATGAGCAACTCACTGTTCGTGCTTTATTTCGGTCTGAAC 3779

Qy       4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686

Db       3780 CATCACCATACGCAATTAGCCCATCATACAGTTTGCTTTGGTCCACGCTATAAGGAACTC 3839

Qy       4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746

Db       3840 ATCGAAGATATTTTTTATCATGACCGACTGTCAGAGGATTTTTCACTCTATCTCCATGCC 3899

Qy       4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806

Db       3900 CCCTCTGTTACTGACCCATCCCTTGCTCCCCGAGGATGCGCAAGTTACTACGTTCTGGCC 3959

Qy       4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866

Db       3960 CCTGTACCACATTTGGGGACAGCTAATTTAAACTGGGATATAGAGGGGCCACGCTTACGT 4019

Qy       4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926

Db       4020 GACCGTATTTTCGCGTATCTGGAAAAGTACTATATGCCCGGTCTGCTCAAGCAGTTAGTC 4079

Qy       4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986

Db       4080 GTTCATCGTATTTTTACCCCATTTGATTTTCGTGACCAACTCAATGCCCATCTCGGTTCG 4139

Qy       4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046

Db       4140 GCCTTCTCCTTTGAACCGTTGCTGACACAAAGCGCTTGGTTCCGACCACACAATCGGGAC 4199

Qy       5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106

Db       4200 AACCGCATTGATAATCTGTATCTCGTCGGCGCAGGCACGCATCCTGGAGCAGGTATTCCC 4259

Qy       5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166

Db       4260 GGCGTGATTGGATCAGCAAAAGCCACCGCCACATTGATGTTAGAGGATATCGCTAAATGA 4319

Qy       5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226

Db       4320 ATCCTGCGC---TACTTAAACACGTTACTCAGATAATGGAGCAAGGCTCAAAAAGTTTTG 4376

Qy       5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286

Db       4377 CCAGTGTCACCCGACTTTTCGATACAGCAACACGACACAGCACAATGATGCTGTACGCCT 4436

Qy       5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346

Db       4437 GGTGCCGTTATTGTGATGATATAATAGACGGGCAAGAGCTGGGGAGACAAATAAGCAGTG 4496
```

```
Qy       5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
               ||      |   ||| |  |      | ||    ||||||      ||    |  || |
Db       4497 TTGATAAGTATAGCGCCCGCGAGAAACTTCAAATGCTGCAATACCTGACAAAGCAGGCTT 4556

Qy       5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
              ||  |||   ||  |||   ||||| || |||   || ||||||  ||| ||  | | |
Db       4557 ACGACGGTTTGCCGATGACCGAACCTGCTTTTGCAGCCTTTCAGACAGTAGCTCTGAGTA 4616

Qy       5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
              |  |  | || || |||| ||| || || ||||||||||| ||||| ||||||| || |
Db       4617 ATGAAATCCCTCAGCAACAGGCCTTCGAACATCTGGAAGGATTCGCAATGGATGTACTCT 4676

Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
                || | ||||   |  || ||||| || ||| | || ||||| || |||||||| || |
Db       4677 GTGAGCCTTACAGAACGTTGGATGATACTCTGAGATATTGCTACCATGTCGCGGGAGTAG 4736

Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
              | ||  ||||||||||  | || ||||| || ||||| |     ||||| ||||  ||||
Db       4737 TTGGGCTGATGATGGCAAGGGTTATGGGGGTTCGCGAGGCGAGTGTGCTGGATCGGGCCT 4796

Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
              ||||| ||||| | || || ||| | ||||||||||| || || ||| | || |||||
Db       4797 GCGATCTAGGAATTGCTTTTCAGTTAACTAACATTGCTCGGGATATTATTGAGGATGCGA 4856

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
              |    || || |||||||| |||| |   |||||  | || |  ||  ||    | || |
Db       4857 AGGCCGGGCGTTGCTATCTTCCGCTAGAATGGCTTCACCAAGAAGGGCTAATGCCAGACA 4916

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
              | || | |    |  || | ||| | ||||| ||     ||||| | || ||  | | ||
Db       4917 CACTTATTTATACCGAAAACCGTCCTGCGCTAGCTCGGGTGGCATCCCGATTGATTGTGG 4976

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
              | |||||| | ||||||  ||| || |   | ||  |    ||||||||  |||| || |
Db       4977 AAGCGGAATCTTATTATACCTCCGCTCTGACAGGGCTTGTTGGTTTACCTTTGCGTTCAG 5036

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | |||| ||||||| | ||||  ||  | ||||| ||||| ||| | ||||| || || |
Db       5037 CCTGGGTCATCGCTTCTGCTCATGGTATCTATCGTGAAATCGGCATTAAAGTGCAACAAG 5096

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              | || || |  |||||||| |   | ||  | || | |   || || ||  |  |  |
Db       5097 CAGGAGTACGGGCCTGGGAATGCAGACAAAGAACTAATCGTGGAGAGAAGGTTACCTTAT 5156

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              |  ||   || |||     |||| | |  || || ||| |     |    || ||
Db       5157 TAATGGCTGGCGCAATGAAGGCGCTTATATCTCGGGTGGCGATAACAACGCCTCGCGATC 5216

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGA 6092
              |     |||||||||| |||| || |
Db       5217 CCAAATTGTGGCAGCGACCTCATTAA 5242

RESULT 13
AAV40146/c
ID   AAV40146 standard; DNA; 8625 BP.
XX
AC   AAV40146;
XX
DT   10-AUG-1999  (first entry)
XX
DE   Flavobacterium sp. R1534 nucleotide sequence.
XX
KW   Carotenoid; pigment; canthaxanthin; R1534; crtB; prephytoene synthase;
KW   crtI; phytoene desaturase; crtY; lycopene cyclase; GGPP synthase; crtE;
KW   crtW2396; beta-carotene beta-oxygenase; food product; fermentation; ds.
XX
OS   Flavobacterium sp.
XX
PN   JP10155497-A.
XX
PD   16-JUN-1998.
XX
PF   02-DEC-1997;   97JP-00348653.
XX
PR   02-DEC-1996;   96EP-00810839.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
DR   WPI; 1998-391048/34.
DR   P-PSDB; AAW69530, AAW69531, AAW69532, AAW69533, AAW69534.
XX
PT   Preparation of carotenoid - comprises fermentation with transformed cell.
XX
PS   Claim 1; Fig 7-21; 80pp; Japanese.
XX
CC   The invention describes the preparation of carotenoid pigments e.g.
CC   canthaxanthins using a cell transformed by a vector having DNA sequences
CC   (a) to (e) or substantially homologous sequences. (a) a DNA sequence
CC   (crtE) coding GGPP synthase of Flavobacterium sp. R1534; (b) a DNA
CC   sequence (crtB) coding prephytoene synthase of Flavobacterium sp. R1534;
CC   (c) a DNA sequence (crtI) coding phytoene desaturase of Flavobacterium
CC   sp. R1534; (d) a DNA sequence (crtY) coding lycopene cyclase of
CC   Flavobacterium sp. R1534, and (e) a DNA sequence (crtW2396) coding beta-
CC   carotene beta-oxygenase of a microbe E-396 (FERM BP-4283). The carotenoid
CC   or a carotenoid mixture can also be used in preparation of food products.
CC   The method is an improved method of fermentation for carotenoid
CC   production
XX
SQ   Sequence 8625 BP; 1458 A; 2898 C; 2964 G; 1295 T; 0 U; 10 Other;

  Query Match             14.8%;  Score 1037;  DB 2;  Length 8625;
  Best Local Similarity   56.6%;  Pred. No. 1.2e-263;
  Matches 2038;  Conservative    0;  Mismatches 1535;  Indels   30;  Gaps    5;

Qy       2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
              | ||  || ||     |  |  |  | ||||    |     || |    | ||| || |
Db       6984 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 6925

Qy       2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
               | |  || || || ||| |   ||   |||| |||||     | ||  |       ||
Db       6924 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 6865

Qy       2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
               |   |     | || |||| ||        |    ||  | | | | || ||||||||
Db       6864 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 6805

Qy       2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
              |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||
Db       6804 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 6745

Qy       2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                 ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db       6744 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 6685

Qy       2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
               | || ||||   | ||||   |   | | | |  ||  |         ||||| |||||||
Db       6684 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 6634

Qy       2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
               ||   || ||    ||      | |  |    ||    | ||||||| |||| |||
Db       6633 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 6574

Qy       2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                  | |  ||       ||||| || |  || ||         ||| || || |||
Db       6573 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 6514

Qy       2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
               ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db       6513 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 6454

Qy       3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
              |  |||||  | ||||| || |||||    |||   |  || || || ||  ||||   |
Db       6453 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 6394

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145
```

```
          |||||  ||||  ||   ||  |  |||| || ||||| |  |||    ||     ||
Db   6393 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 6334

Qy   3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205
          ||||  ||   | |       |||         |||||||||     ||  ||||||||
Db   6333 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6274

Qy   3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265
            |     | ||| ||| |||  |  |  | |                   | |  |||
Db   6273 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6214

Qy   3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325
          |     |   |  | ||   ||  ||    | |      || ||| | || ||||||||
Db   6213 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6154

Qy   3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385
          ||  ||  ||||||| || ||  ||||  | |||| |  | | |||||     | || |||
Db   6153 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6094

Qy   3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445
          |     |  ||    ||   | | |   ||        ||   |      |  ||    |
Db   6093 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6034

Qy   3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505
             ||||   | || ||  | || ||||| ||||| ||||||||    | | |   |||
Db   6033 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 5977

Qy   3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565
           | |  ||||  ||| || ||||| || || ||||| ||         | | ||    |||
Db   5976 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 5917

Qy   3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625
          || || |||||||  |||    ||  |||       || ||  |   |||    |||||
Db   5916 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 5857

Qy   3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685
           | ||  |     | ||  |||||   |||    |      |   |    | ||| |  |
Db   5856 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 5797

Qy   3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745
          |||    | |    || || |||||||| || ||||| || ||||| || || |||||||
Db   5796 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 5737

Qy   3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805
          || |||| ||||||  | |||||| |  |||||   |||||||| || ||| || |||||||
Db   5736 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 5677

Qy   3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865
          |||| |   | ||    |||     ||||||||| || ||  |||||  | ||||| |||
Db   5676 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 5617

Qy   3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925
           |  | |   |||| |||   | ||    ||  || || |      | ||||   ||||
Db   5616 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 5557

Qy   3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985
          |||||||  |||||||||| || |||   ||||  || ||||       | || ||||  |
Db   5556 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 5497

Qy   3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045
          |  |  |      ||||   |  ||| |||| || | ||||||||    |||    ||| || ||||
Db   5496 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 5437

Qy   4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105
          |||| |  || |  |  ||  |    || || | | | ||||| |||||||| || || || |
Db   5436 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 5377

Qy   4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165
          |   ||| || ||| || || |  | ||||||  || |||||||  | ||   || ||||
Db   5376 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5317

Qy   4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225
          | ||||  |||   ||| ||||||||||||||   || || ||||||   |||||||| |
Db   5316 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5257

Qy   4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285
          |||| || || |||||| | ||||||| ||||| || || ||| | || ||   |||||||
Db   5256 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5197

Qy   4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345
          || |  | || ||||||||||| ||      | ||||| ||||||  |    |||||||||
Db   5196 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5137

Qy   4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405
           |  ||||||   ||||||||     |||||||||    | ||||||  |||| | ||||
Db   5136 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5077

Qy   4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465
          |||||   || |  | | | || |||   ||| |||||   |  |||||   | |   ||
Db   5076 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5017

Qy   4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525
           ||| || |  ||    || |   | ||||||  ||| | ||||||| |||| | |||
Db   5016 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 4957

Qy   4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
           | | |||||  ||||    |  |  |     |     ||  |    ||||| || |||
Db   4956 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 4897

Qy   4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645
          ||  |   |  |||||||| || ||  ||||| ||||||    |||  |  ||  | || |
Db   4896 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 4837

Qy   4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705
          | |||||||  ||||   || |||||||| ||    | | || |  | ||||| ||||||
Db   4836 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 4777

Qy   4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765
          |    |  ||| | |||||||| ||  | | |||||  ||||||||   | ||||||
Db   4776 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 4717

Qy   4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825
            ||||| ||    ||      ||    ||| ||  | ||||| || |||||||| |||
Db   4716 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 4657

Qy   4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885
           ||| ||  |||| ||| ||   || |  ||| |||    ||| || || || |   | |
Db   4656 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 4597

Qy   4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
          ||||||  | |  ||| || |   | ||  ||||  |         | | || || |  |
Db   4596 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 4537

Qy   4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
          |    |||||  ||  |   |||   ||||||||||||     || || ||| | |||||||
Db   4536 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 4477

Qy   5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
          |  ||||||||  ||||||||||||||||||||||||||||   ||  |||   ||  |||
Db   4476 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 4417

Qy   5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
          |  | ||||| || || |||||||| || |||||| | || ||  | ||    ||||
Db   4416 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 4357

Qy   5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
          ||||||| ||   | | ||||||  ||| |  ||||    |      | |
Db   4356 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4297

Qy   5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
              | | ||    |||| |||  ||| ||| | | || ||||    || ||||||||| |
Db   4296 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4240

Qy   5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
               |    |  |||      ||| | |||||| ||||| |||||  | ||    |||||
Db   4239 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4180
```

```
Qy          5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
                  ||||| || |||||   | ||||    | |  |  |      || | ||  ||| | ||
Db          4179 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4126

Qy          5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
                 ||| ||  ||  |   ||||     || || |    |||      ||  |    |  |
Db          4125 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4066

Qy          5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
                     ||| | ||   |||| | |   | || ||       ||  |  | || | |
Db          4065 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4006

Qy          5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
                 | || | || |   | || || ||||| |||||||      |     | |||   ||| |
Db          4005 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 3946

Qy          5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
                  ||||||  ||||   || | ||| ||||||||||| || || ||  |||||||||||||
Db          3945 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 3886

Qy          5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
                  || ||||||||||  ||||| |||||||| |||| |||||||||| | || || |||||
Db          3885 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 3826

Qy          5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
                 |||||| || ||||| || |||||| |  | || |||||||  |  || |||||||||||
Db          3825 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 3766

Qy          5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
                 |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db          3765 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 3715

Qy          5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
                 |   |  ||||||   ||  ||      ||  |  | |  | ||| || |||||||||
Db          3714 TTCGGACGCGCTCTATTCCGTGATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 3655

Qy          5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
                 |||||||||    ||  | |||  | | |||| |||||  |||||| | |||||  | ||
Db          3654 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 3595

Qy          5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
                  |   |  | ||||||| ||| ||        | | ||| |  ||     | ||||
Db          3594 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 3535

Qy          5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                     |||    ||||    || | ||  ||  | | ||| |||| |   || |
Db          3534 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 3475

Qy          6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
                  |||  | | |||||  ||    |    |||    | |  |  || ||||||   || ||
Db          3474 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 3415

Qy          6086 TCG 6088
                  ||
Db          3414 GCG 3412

RESULT 14
AAV40151/c
ID   AAV40151 standard; DNA; 11233 BP.
XX
AC   AAV40151;
XX
DT   10-AUG-1999  (first entry)
XX
DE   DNA sequence of plasmid pZea4.
XX
KW   Carotenoid; pigment; canthaxanthin; R1534; crtB; prephytoene synthase;
KW   crtI; phytoene desaturase; crtY; lycopene cyclase; crtW2396; pZea4;
KW   beta-carotene beta-oxygenase; food product; fermentation; ds.
XX
OS   Synthetic.
XX
PN   JP10155497-A.
XX
PD   16-JUN-1998.
XX
PF   02-DEC-1997;   97JP-00348653.
XX
PR   02-DEC-1996;   96EP-00810839.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
DR   WPI; 1998-391048/34.
XX
PT   Preparation of carotenoid - comprises fermentation with transformed cell.
XX
PS   Example 2; Fig 42-53; 80pp; Japanese.
XX
CC   The invention describes the preparation of carotenoid pigments e.g.
CC   canthaxanthins using a cell transformed by a vector having DNA sequences
CC   (a) to (e) or substantially homologous sequences. (a) a DNA sequence
CC   (crtE) coding GGPP synthase of Flavobacterium sp. R1534; (b) a DNA
CC   sequence (crtB) coding prephytoene synthase of Flavobacterium sp. R1534;
CC   (c) a DNA sequence (crtI) coding phytoene desaturase of Flavobacterium
CC   sp. R1534; (d) a DNA sequence (crtY) coding lycopene cyclase of
CC   Flavobacterium sp. R1534, and (e) a DNA sequence (crtW2396) coding beta-
CC   carotene beta-oxygenase of a microbe E-396 (FERM BP-4283). The carotenoid
CC   or a carotenoid mixture can also be used in preparation of food products.
CC   The method is an improved method of fermentation for carotenoid
CC   production
XX
SQ   Sequence 11233 BP; 2142 A; 3525 C; 3605 G; 1960 T; 0 U; 1 Other;

  Query Match             14.8%;  Score 1037;  DB 2;  Length 11233;
  Best Local Similarity   56.6%;  Pred. No. 1.4e-263;
  Matches 2038;  Conservative    0;  Mismatches 1535;  Indels   30;  Gaps    5;

Qy          2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
                 | ||  || ||     |  |  |  | ||||    |     || |    | ||| || |
Db          7666 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 7607

Qy          2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
                  | |  || || || ||| |   ||   |||| |||||     | ||  |       ||
Db          7606 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 7547

Qy          2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
                  |   |     | || |||| ||        |    ||  | | | | || ||||||||
Db          7546 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 7487

Qy          2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
                 |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||
Db          7486 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 7427

Qy          2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                    ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db          7426 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 7367

Qy          2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
                  | || ||||   | ||||   |   | | |  ||  |         ||||| |||||||
Db          7366 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 7316

Qy          2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
                  ||   || ||    ||      | |  |    ||    | ||||||| |||| |||
Db          7315 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 7256

Qy          2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                     | |  ||       ||||| || |  || ||         ||| || || |||
Db          7255 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 7196

Qy          2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
                  ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db          7195 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 7136

Qy          3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
```

```
Db       7135 CGCCCGATGATCATGGACGCGACCGTTGCGCAGATGGACGGGTACCGCTTCATCTATCTG 7076

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145

Db       7075 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 7016

Qy       3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205

Db       7015 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6956

Qy       3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265

Db       6955 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6896

Qy       3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325

Db       6895 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6836

Qy       3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385

Db       6835 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6776

Qy       3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445

Db       6775 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6716

Qy       3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505

Db       6715 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 6659

Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565

Db       6658 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 6599

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625

Db       6598 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 6539

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685

Db       6538 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 6479

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745

Db       6478 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 6419

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805

Db       6418 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 6359

Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865

Db       6358 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 6299

Qy       3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925

Db       6298 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 6239

Qy       3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985

Db       6238 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 6179

Qy       3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045

Db       6178 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 6119

Qy       4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105

Db       6118 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 6059

Qy       4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGACGTCTGC 4165

Db       6058 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5999

Qy       4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225

Db       5998 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5939

Qy       4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285

Db       5938 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5879

Qy       4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345

Db       5878 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5819

Qy       4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405

Db       5818 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5759

Qy       4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465

Db       5758 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5699

Qy       4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525

Db       5698 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 5639

Qy       4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585

Db       5638 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 5579

Qy       4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645

Db       5578 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 5519

Qy       4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705

Db       5518 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 5459

Qy       4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765

Db       5458 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 5399

Qy       4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825

Db       5398 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 5339

Qy       4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885

Db       5338 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 5279

Qy       4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945

Db       5278 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 5219

Qy       4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005

Db       5218 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 5159

Qy       5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065

Db       5158 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 5099

Qy       5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125

Db       5098 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 5039

Qy       5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185

Db       5038 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4979

Qy       5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245

Db       4978 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4922
```

```
Qy          5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
                      |    |  |||      ||| | ||||| ||||| |||||  | ||    |||||
Db          4921 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4862

Qy          5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
                  ||||| || |||||   | ||||    | |  |  |      || | ||   || | ||
Db          4861 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4808

Qy          5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
                 ||| ||  ||  |   ||||     || || |    |||      ||  |    |  |
Db          4807 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4748

Qy          5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
                     ||| | ||   |||| | |   | || ||       ||  |  | ||| | |
Db          4747 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4688

Qy          5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
                 | || | || |   | || || ||||| |||||||      |     | |||   ||| |
Db          4687 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 4628

Qy          5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
                  ||||||  ||||   || | ||| ||||||||||| || || ||  |||||||||||||
Db          4627 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 4568

Qy          5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
                  || ||||||||||  ||||| |||||||| |||| |||||||||| | || || |||||
Db          4567 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 4508

Qy          5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
                 |||||| || ||||| || |||||| |  | || |||||||  |  || |||||||||||
Db          4507 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 4448

Qy          5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
                 |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db          4447 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 4397

Qy          5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
                 |   |  ||||||   ||  |       ||  |  | |  | ||| || |||||||||
Db          4396 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 4337

Qy          5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
                 |||||||||    ||  | |||  | | |||| |||||  |||||| | |||||  | ||
Db          4336 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 4277

Qy          5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
                  |   |  | ||||||| ||| ||        | | ||| |  ||     | ||||
Db          4276 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 4217

Qy          5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                     |||    ||||    || | ||  ||  | | ||| |||| |   || |
Db          4216 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 4157

Qy          6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
                  |||  | | |||||  ||    |    |||    | |  |  || ||||||   || ||
Db          4156 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 4097

Qy          6086 TCG 6088
                  ||
Db          4096 GCG 4094


RESULT 15
AAT45143/c
ID   AAT45143 standard; DNA; 8625 BP.
XX
AC   AAT45143;
XX
DT   17-OCT-2003  (revised)
DT   08-MAR-1997  (first entry)
XX
DE   Flavobacterium carotenoid biosynthesis cluster DNA.
XX
```

```
KW   Carotenoid; lycopene; beta-carotene; echinenone; canthaxanthin;
KW   zeaxanthin; adonixanthin; astaxanthin; crtE;
KW   geranylgeranyl pyrophosphate synthase; GGPP synthase; crtB;
KW   prephytoene synthase; phytoene synthase; crtI; phytoene desaturase; crtY;
KW   lycopene cyclase; crtZ; beta-carotene hydroxylase; ds.
XX
OS   Flavobacterium sp. ATCC 21588; WT (ATCC 21588).
XX
FH   Key             Location/Qualifiers
FT   CDS             2. .1168
FT                   /*tag=  a
FT                   /label= ORF-5
FT   CDS             1180. .2355
FT                   /*tag=  b
FT                   /label= ORF-1
FT   CDS             2521. .3408
FT                   /*tag=  c
FT                   /label= CrtE
FT                   /product= "GGPP synthase"
FT   CDS             complement(3405. .4316)
FT                   /*tag=  d
FT                   /label= CrtB
FT                   /product= "(pre)phytoene synthase"
FT   CDS             complement(4313. .5797)
FT                   /*tag=  e
FT                   /label= CrtI
FT                   /product= "phytoene desaturase"
FT   CDS             complement(5794. .6942)
FT                   /*tag=  f
FT                   /label= CrtY
FT                   /product= "lycopene cyclase"
FT   CDS             complement(6939. .7448)
FT                   /*tag=  g
FT                   /label= CryZ
FT                   /product= "beta-carotene hydroxylase"
FT   CDS             complement(7767. .8315)
FT                   /*tag=  h
FT                   /label= ORF-16
FT   misc_difference 8348. .8349
FT                   /*tag=  i
FT                   /note= "bases 8348-8349 are given as nn in the
FT                   specification"
FT   misc_difference 8539. .8540
FT                   /*tag=  j
FT                   /note= "bases 8539-8540 are given as nn in the
FT                   specification"
FT   misc_difference 8581
FT                   /*tag=  k
FT                   /note= "base 8581 is given as n in the specification"
FT   misc_difference 8590
FT                   /*tag=  l
FT                   /note= "base 8590 is given as n in the specification"
FT   misc_difference 8592
FT                   /*tag=  m
FT                   /note= "base 8592 is given as n in the specification"
FT   misc_difference 8602. .8604
FT                   /*tag=  n
FT                   /note= "bases 8602-8604 are given as nnn in the
FT                   specification"
XX
PN   EP747483-A2.
XX
PD   11-DEC-1996.
XX
PF   29-MAY-1996;   96EP-00108556.
XX
PR   09-JUN-1995;   95EP-00108888.
XX
PA   (HOFF ) HOFFMANN LA ROCHE & CO AG F.
XX
PI   Hohmann H,  Pasamontes L,  Tessier M,  Van Loon A;
XX
DR   WPI; 1997-023160/03.
DR   P-PSDB; AAW06513, AAW06514, AAW06515, AAW06516, AAW00871, AAW06517,
DR   AAW06518, AAW06519.
XX
PT   Flavobacterium gene sequences encoding carotenoid biosynthesis enzymes -
PT   for the production of carotenoid(s), useful in foods and animal feeds.
XX
PS   Example 2; Fig 7; 80pp; English.
XX
CC   Genomic DNA (AAT45143) of Flavobacterium sp. R1534 includes genes of the
CC   carotenoid biosynthesis pathway. The sequence was deduced from inserts of
CC   6 clones obtd. from genomic libraries e.g. by PCR amplification (see also
CC   AAT45144-45) and use of partial clones to screen the library. The
CC   identities of the gene products (see also AAW06513-19 and AAW00871) were
CC   detd. by examining carotenoid accumulation in E. coli hosts transformed
CC   with deleted variants of the gene cluster. The isolated genes can be used
CC   in different combinations to produce carotenoids in transformed host
CC   cells. (Updated on 17-OCT-2003 to standardise OS field)
XX
SQ   Sequence 8625 BP; 1458 A; 2898 C; 2964 G; 1295 T; 0 U; 10 Other;

  Query Match             14.8%;  Score 1035.4;  DB 2;  Length 8625;
  Best Local Similarity   56.5%;  Pred. No. 3.3e-263;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy       2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
              | ||  || ||     |  |  |  | ||||    |     || |    | ||| || |
Db       6984 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 6925

Qy       2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
               | |  || || || ||| |   ||   |||| |||||     | ||  |       ||
Db       6924 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 6865

Qy       2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
               |   |     | || |||| ||        |    ||  | | | | || ||||||||
Db       6864 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 6805

Qy       2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
              |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||
Db       6804 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 6745

Qy       2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                 ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db       6744 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 6685

Qy       2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
               | || ||||   | ||||   |   | | |  ||  |         ||||| |||||||
Db       6684 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 6634

Qy       2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
               ||   || ||    ||      | |  |    ||    | ||||||| |||| |||
Db       6633 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 6574

Qy       2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                  | |  ||       ||||| || |  || ||         ||| || || |||
Db       6573 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 6514

Qy       2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
               ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db       6513 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 6454

Qy       3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
              |  |||||  | ||||| || |||||    |||   |  || || || ||  ||||   |
Db       6453 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 6394

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145
              |||||  ||||  ||   ||  |  |||| || ||||| |  |||    ||     ||
Db       6393 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 6334

Qy       3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205
              ||||  ||   | |       |||            |||||||||     ||  ||||||||
Db       6333 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6274

Qy       3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265
                |     | ||| ||| |||  |  |  |  |                    | |  |||
Db       6273 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6214

Qy       3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325
              |     |   |  | ||   ||  ||    | |      || ||| | || |||||||||
Db       6213 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6154

Qy       3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385
              ||  ||  ||||||| || ||  |||| |  |||| |  | | |||||     | || |||
Db       6153 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6094

Qy       3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445
              |     |  ||    ||   | | |   ||        ||   |      |  ||    |
Db       6093 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6034

Qy       3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505
                 ||||   | || ||  |  || ||||| ||||| ||||||||    | | |   |||
Db       6033 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 5977

Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565
               | |  ||||  ||| || ||| || || ||||| ||         | | ||    |||
Db       5976 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 5917

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625
              || || |||||||  |||    ||  |||         || ||  |   |||     |||||
Db       5916 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 5857

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685
               | ||  |     | ||  |||||   |||    |      |   |    | ||| |  |
Db       5856 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 5797

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745
              |||    | |    || || |||||||| || ||||| || ||||| || || |||||||
Db       5796 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 5737

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805
              || |||| |||||  | |||||| |  |||||   ||||||||| || ||||| |||||||
Db       5736 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 5677

Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865
              |||| |   | ||    |||     ||||||||| || || |||||  | ||||| |||
Db       5676 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 5617

Qy       3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925
               |  | |   |||| |||   | ||    ||  || || |      | ||||   ||||
Db       5616 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 5557

Qy       3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985
              |||||||  |||||||||| || |||   ||||  || |||       | || ||||  |
Db       5556 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 5497

Qy       3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045
              |  |  |      ||||   |  ||| ||||| | |||||||||    |||||| || ||||
Db       5496 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 5437

Qy       4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105
              |||| |  || |  |  ||  |    || || | | | ||||| |||||||| || || |
Db       5436 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 5377

Qy       4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165
              |   ||| || ||| || || | |  | ||||||  || |||||||  | ||   || ||||
Db       5376 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5317

Qy       4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225
              | ||||  |||   ||| ||||||||||||||   || || ||||||   |||||||| |
Db       5316 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5257

Qy       4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285
              |||| || || |||||| ||||||| ||||| || ||  ||| || ||| | ||    |||||||
Db       5256 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5197

Qy       4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345
              || |  | || ||||||||||| ||     | ||||| |||||  |    |||||||||
Db       5196 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5137

Qy       4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405
               |  ||||||   |||||||||     |||||||    | ||||||  ||||| | ||||
Db       5136 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5077

Qy       4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465
              |||||   || |  | | | || |||   ||| |||||   |  |||||   | |   ||
Db       5076 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5017

Qy       4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525
               ||| || |  ||    || |   | ||||||  ||| | ||||||| |||| | |||
Db       5016 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 4957

Qy       4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
               | | |||||  ||||    |  |  |     |     ||  |    |||||| ||  |||
Db       4956 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 4897

Qy       4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645
              ||  |   |  |||||||| || ||  ||||| ||||||    |||  |  ||  | || |
Db       4896 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 4837

Qy       4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705
              | ||||||| ||||   || |||||||| ||    | | || |  | |||||| ||||||
Db       4836 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 4777

Qy       4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765
              |    |  ||| | |||||||| ||  | | | |||||  ||||||||||   | ||||||
Db       4776 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 4717

Qy       4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825
                ||||| ||    ||      ||    ||| ||  | ||||| || |||||||| |||
Db       4716 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 4657

Qy       4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885
               ||| ||  |||| ||| ||   || |  ||| |||    ||| || || || |   | |
Db       4656 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 4597

Qy       4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
              ||||||  | |  ||| || |   | ||  ||||  |         | | || || |  |
Db       4596 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 4537

Qy       4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
              |    |||||  ||  |   |||   ||||||||||||   || || ||| | |||||||
Db       4536 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 4477

Qy       5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
              |  |||||||||  ||||||||||||||||||||||||||||   ||  |||   ||  |||
Db       4476 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 4417

Qy       5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
              |  | ||||| || || ||||||| || ||||| | || ||  | || ||    ||||
Db       4416 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 4357

Qy       5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
              ||||||| ||   | | ||||||  ||| |  ||||    |      | |
Db       4356 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4297

Qy       5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
                  | | ||    |||| |||  ||| ||| | || |||||    || |||||||| |
Db       4296 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4240

Qy       5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
                   |    |  |||      ||| | ||||| ||||| |||||  | ||    |||||
Db       4239 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4180

Qy       5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
               ||||| || |||||   | ||||    | |  |  |  |      || | ||  ||| | ||
Db       4179 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4126

Qy       5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
              ||| ||  ||  |   ||||     || || |    |||      ||  |    |  |
Db       4125 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4066
```

```
Qy       5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
                  ||| | ||   |||| | |   | || ||        ||  |  | || | |
Db       4065 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4006

Qy       5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
              | || | || |   | || || ||||| |||||||      |     | |||   ||| |
Db       4005 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 3946

Qy       5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
               ||||||  ||||   || | ||| ||||||||||| || || ||  |||||||||||||
Db       3945 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 3886

Qy       5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
               || ||||||||||  ||||| |||||||| |||| ||||||||||| | || || |||||
Db       3885 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 3826

Qy       5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
              |||||| || ||||| || |||||| |  | || ||||||||  |  || |||||||||||
Db       3825 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 3766

Qy       5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
              |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db       3765 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 3715

Qy       5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
              |   |  ||||||   ||  |       ||  |  | |  | ||| || |||||||||
Db       3714 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 3655

Qy       5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
              |||||||||    ||  | ||  | | |||| |||||  |||||| | |||||  | ||
Db       3654 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 3595

Qy       5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
               |   |  | ||||||| ||| ||          | | ||| |  ||      | ||||
Db       3594 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 3535

Qy       5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                  |||     ||||    || | ||  ||  | | |||  |||| |   | |
Db       3534 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 3475

Qy       6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
               |||  | | |||||  ||    |    |||    | |  |  || ||||||   || ||
Db       3474 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 3415

Qy       6086 TCG 6088
               ||
Db       3414 GCG 3412


Search completed: January 25, 2005, 05:06:30
Job time : 2953 secs
```

```
                  GenCore version 5.1.6
           Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 25, 2005, 03:38:11 ; Search time 501 Seconds
                                          (without alignments)
                                          9929.763 Million cell updates/sec

Title:          US-10-808-979-18
Perfect score:  6999
Sequence:       1 gcgatggcaatggaaaatgt..........gacgttttagacatgcaata 6999

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters:      1649014

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/1/ina/backfiles1.seq:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                SUMMARIES
                    %
Result             Query
  No.     Score    Match Length DB  ID                        Description
 -----------------------------------------------------------------------------
     1   2624.6    37.5    6918  1  US-07-783-705A-13          Sequence 13, Appl
c    2   1035.4    14.8    8625  3  US-08-980-832-1            Sequence 1, Appli
c    3   1035.4    14.8    8625  4  US-09-920-923B-1           Sequence 1, Appli
c    4   1035.4    14.8   11233  3  US-08-980-832-27           Sequence 27, Appl
c    5   1035.4    14.8   11233  4  US-09-920-923B-27          Sequence 27, Appl
     6    826.2    11.8    1479  1  US-07-783-705A-10          Sequence 10, Appl
     7    713.4    10.2    1518  1  US-08-095-726-7            Sequence 7, Appli
     8    713.4    10.2    1518  1  US-08-096-043-7            Sequence 7, Appli
     9    713.4    10.2    1518  1  US-08-096-623A-7           Sequence 7, Appli
    10    708.4    10.1    1522  1  US-08-095-726-9            Sequence 9, Appli
    11    708.4    10.1    1522  1  US-08-096-043-9            Sequence 9, Appli
    12    708.4    10.1    1522  1  US-08-096-623A-9           Sequence 9, Appli
    13    619.4     8.8    1482  3  US-08-660-645A-6           Sequence 6, Appli
    14    619.4     8.8    1482  3  US-09-298-718-6            Sequence 6, Appli
    15    619.4     8.8    1482  3  US-09-546-969-6            Sequence 6, Appli
    16    619.4     8.8    1482  4  US-09-547-267-6            Sequence 6, Appli
    17    515.6     7.4    1198  1  US-08-095-726-5            Sequence 5, Appli
    18    515.6     7.4    1198  1  US-08-096-043-5            Sequence 5, Appli
    19    515.6     7.4    1198  1  US-08-093-577-5            Sequence 5, Appli
    20    515.6     7.4    1198  1  US-08-096-623A-5           Sequence 5, Appli
    21    449.6     6.4    1083  1  US-08-331-004A-1           Sequence 1, Appli
    22    449.6     6.4    1083  5  PCT-US95-13937A-1          Sequence 1, Appli
    23      438     6.3    1235  1  US-08-095-726-13           Sequence 13, Appl
    24      438     6.3    1235  1  US-08-096-623A-13          Sequence 13, Appl
    25      430     6.1    1232  4  US-08-908-758-1            Sequence 1, Appli
    26      430     6.1    1232  4  US-09-023-587A-1           Sequence 1, Appli
    27    426.8     6.1    1235  1  US-08-095-726-15           Sequence 15, Appl
    28    426.8     6.1    1235  1  US-08-096-623A-15          Sequence 15, Appl
    29    419.8     6.0    1296  1  US-07-783-705A-8           Sequence 8, Appli
    30      414     5.9     891  1  US-07-783-705A-11          Sequence 11, Appl
c   31    399.6     5.7     947  1  US-08-096-623A-16          Sequence 16, Appl
c   32    399.6     5.7     947  1  US-08-096-623A-17          Sequence 17, Appl
    33    386.2     5.5    1149  1  US-07-783-705A-9           Sequence 9, Appli
    34    357.2     5.1    2886  1  US-08-663-310-7            Sequence 7, Appli
    35    357.2     5.1    2886  2  US-09-006-491-7            Sequence 7, Appli
    36    357.2     5.1    2886  3  US-09-335-919-7            Sequence 7, Appli
    37    346.2     4.9     909  1  US-07-783-705A-7           Sequence 7, Appli
    38    308.2     4.4    1157  1  US-08-095-726-1            Sequence 1, Appli
    39    308.2     4.4    1157  1  US-08-095-726-3            Sequence 3, Appli
    40    308.2     4.4    1157  1  US-08-096-043-1            Sequence 1, Appli
    41    308.2     4.4    1157  1  US-08-096-043-3            Sequence 3, Appli
    42    308.2     4.4    1157  1  US-08-093-577-1            Sequence 1, Appli
    43    308.2     4.4    1157  1  US-08-093-577-3            Sequence 3, Appli
    44    308.2     4.4    1157  1  US-08-096-623A-1           Sequence 1, Appli
    45    308.2     4.4    1157  1  US-08-096-623A-3           Sequence 3, Appli

                               ALIGNMENTS

RESULT 1
US-07-783-705A-13
; Sequence 13, Application US/07783705A
; Patent No. 5429939
;  GENERAL INFORMATION:
;    APPLICANT:  Misawa, No. 5429939ihiko
;    APPLICANT:  Kobayashi, Kazuo
;    APPLICANT:  Nakamura, Katsumi
;    APPLICANT:  Yamano, Shigeyuki
;    TITLE OF INVENTION:  DNA SEQUENCES USEFUL FOR THE
;    TITLE OF INVENTION:  SYNTHESIS OF CAROTENOIDS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ladas & Parry
;      STREET:  26 West 61 Street
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720Kb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  N/A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/783,705A
;      FILING DATE:  19911023
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-103078
;      FILING DATE:  21-APR-1989
;      APPLICATION NUMBER:  JP 2-53225
;      FILING DATE:  05-MAR-1990
;      APPLICATION NUMBER:  US 07/519,011
;      FILING DATE:  19-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schwadron, Janet I.
;      REGISTRATION NUMBER:  33,778
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-708-1935
;      TELEFAX:  212-246-5959
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6918 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid (plasmid DNA)
;    ORIGINAL SOURCE:
;      ORGANISM:  Erwinia uredovora
;      STRAIN:  20D3
;      CELL TYPE:  unicellular organism
;    FEATURE:
;      NAME/KEY:  CDS
;      NAME/KEY:  /gene="crt E"
;      LOCATION:  225 to 1133
;      IDENTIFICATION METHOD:  by experiment
;      OTHER INFORMATION:  codes for geranylgeranyl
;      OTHER INFORMATION:  pyrophosphpate synthase
;      NAME/KEY:  CDS
;      NAME/KEY:  /gene="crt X"
;      LOCATION:  1143 to 2438
;      IDENTIFICATION METHOD:  by experiment
;      OTHER INFORMATION:  codes for zeaxanthin glycosilase
;      NAME/KEY:  CDS
;      NAME/KEY:  /gene="crt Y"
;      LOCATION:  2422 to 3570
;      IDENTIFICATION METHOD:  by experiment
;      OTHER INFORMATION:  codes for lycopene cyclase
;      NAME/KEY:  CDS
;      NAME/KEY:  /gene="crt I"
;      LOCATION:  3582 to 5060
;      IDENTIFICATION METHOD:  by experiment
;      OTHER INFORMATION:  codes for phytoene desaturase
;      NAME/KEY:  CDS
;      NAME/KEY:  /gene="crt B"
;      LOCATION:  5096 to 5986
;      IDENTIFICATION METHOD:  by experiment
;      OTHER INFORMATION:  codes for phytoene synthase
;      NAME/KEY:  CDS
;      NAME/KEY:  /gene="crt Z"
;      NAME/KEY:  /note="on the complementary strand"
;      LOCATION:  5925 to 6452
;      IDENTIFICATION METHOD:  by experiment
;      OTHER INFORMATION:  codes for @ @-carotene hydroxylase
;    PUBLICATION INFORMATION:
;      AUTHORS:  Misawa, No. 5429939ihiko
;      AUTHORS:  Nakagawa, Masaya
;      AUTHORS:  Kobayashi, Kazuo
;      AUTHORS:  Yamano, Shigeyuki
;      AUTHORS:  Izawa, Yuko
;      AUTHORS:  Nakamura, Katsumi
;      AUTHORS:  Harashima, Keiji
;      TITLE:  Elucidation of the Erwinia uredovora Carotenoid Biosynthesic Path
;      TITLE:  by Functional Analysis of Gene Products Expressed in Escheric
;      JOURNAL:  Journal of Bacteriology
;      VOLUME:  172
;      ISSUE:  12
;      PAGES:  6704-6712
;      DATE:  DEC-1990
US-07-783-705A-13

  Query Match             37.5%;  Score 2624.6;  DB 1;  Length 6918;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4007;  Conservative    0;  Mismatches 2169;  Indels   21;  Gaps    6;

Qy         400 CTTGAACATTTACTGCCTGCCGGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCC 459
               ||||| || ||| |||| |  |      || | |||   ||| |||||||||||||||
Db         303 CTTGATCAGTTATTGCCCGTGGAGGGAGAACGGGATGTTGTGGGTGCCGCGATGCGTGAA 362

Qy         460 GGAACGCTGGCGCAGGGCAAACGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGAT 519
               ||  ||||||| | ||| ||||||||||| ||  | || ||| ||||| | || ||||||
Db         363 GGTGCGCTGGCACCGGGAAAACGTATTCGCCCCATGTTGCTGTTGCTGACCGCCCGCGAT 422

Qy         520 ATGGGTTGCGAGCTGACGCAAAATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTG 579
                |||||||||   | |  ||  | ||  | || ||| | |||||||| ||||||||||||
Db         423 CTGGGTTGCGCTGTCAGCCATGACGGATTACTGGATTTGGCCTGTGCGGTGGAAATGGTC 482

Qy         580 CACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGT 639
               |||||||| |||||||| || || || || ||||  |||||  | ||| || |||| ||
Db         483 CACGCGGCTTCGCTGATCCTTGACGATATGCCCTGCATGGACGATGCGAAGCTGCGGCGC 542

Qy         640 GGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCG 699
               || ||||||||| | |||    | |  || ||  | ||||| || || || ||  | ||
Db         543 GGACGCCCTACCATTCATTCTCATTACGGAGAGCATGTGGCAATACTGGCGGCGGTTGCC 602

Qy         700 CTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAA 759
                |||| ||    || ||||  || ||||||  ||||   ||  |  |  |       |||
Db         603 TTGCTGAGTAAAGCCTTTGGCGTAATTGCCGATGCAGATGGCCTCACGCCGCTGGCAAAA 662

Qy         760 TCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAA 819
                 |   ||| || ||||||| ||    ||| ||||| |||| || || || || || ||
Db         663 AATCGGGCGGTTTCTGAACTGTCAAACGCCATCGGCATGCAAGGATTGGTTCAGGGTCAG 722

Qy         820 TTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAAC 879
               ||| ||||||||   || ||     ||   |||||| | ||||| ||    ||||| ||
Db         723 TTCAAGGATCTGTCTGAAGGGGATAAGCCGCGCAGCGCTGAAGCTATTTTGATGACGAAT 782

Qy         880 GAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGAC 939
                |  | |||||||||  ||||||| | ||| |  |||| |||||  ||||||  ||  |
Db         783 CACTTTAAAACCAGCACGCTGTTTTGTGCCTCCATGCAGATGGCCTCGATTGTTGCGAAT 842

Qy         940 GCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTT 999
               || ||     | | |||  |  | ||       ||    |  ||| | || ||||| |||
Db         843 GCCTCCAGCGAAGCGCGTGATTGCCTGCATCGTTTTTCACTTGATCTTGGTCAGGCATTT 902

Qy        1000 CAACTGCTCGACGACCTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGAT 1059
               |||||||| |||||  |  |||| ||    |   |||||||||| |||    | |||||
Db         903 CAACTGCTGGACGATTTGACCGATGGCATGACCGACACCGGTAAGGATAGCAATCAGGAC 962

Qy        1060 CAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGC 1119
                  || ||||| ||||||||  |  |  | ||  |    |  |  |||   || ||| |
Db         963 GCCGGTAAATCGACGCTGGTCAATCTGTTAGGCCCGAGGGCGGTTGAAGAACGTCTGAGA 1022

Qy        1120 GATCACCTGCGCAGCGCAGATGCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACT 1179
                | || || |     ||   ||  || ||  |    ||||||||| |  ||   |||||||
Db        1023 CAACATCTTCAGCTTGCCAGTGAGCATCTCTCTGCGGCCTGCCAACACGGGCACGCCACT 1082

Qy        1180 CGCCAATATATGCACGCGCTGTTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTC 1239
               |  || | ||| || ||   |||| |  || | ||  |     |||  | ||   |
Db        1083 CAACATTTTATTCAGGCCTGGTTTGACAAAAAACTCGCTGCCGTCAGTTAAGGATGCTGC 1142

Qy        1240 AGCCGGTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGC 1299
               |     || |||| || |||| |||||| |||||  | ||||||||  ||| ||| || |
Db        1143 A-----TGAGCCATTTCGCGGCGATCGCACCGCCTTTTTACAGCCATGTTCGCGCATTAC 1197

Qy        1300 AGGCGTTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCG 1359
               ||    | || ||    ||| | ||||||||| |||||  | || || || ||||||  ||
Db        1198 AGAATCTCGCTCAGGAACTGGTCGCGCGCGGTCATCGGGTGACCTTTATTCAGCAATACG 1257

Qy        1360 ATGCCCGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGC 1419
               ||         ||| |      ||||  ||| |  |||  | |||||||   | | |  |
Db        1258 ATATTAAACACTTGATCGATAGCGAAACCATTGGATTTCATTCCGTCGGGACAGACAGCC 1317

Qy        1420 ATCCTGCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGC 1479
               ||||  ||||  || |  ||| |||| | || | |  |||    ||    || | |||
Db        1318 ATCCCCCCGGCGCGTTAACGCGCGTGCTACACCTGGCGGCTCATCCTCTGGGGCCGTCAA 1377

Qy        1480 TGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTG 1539
               || |     | ||| |  |    |||  |  |||||||||||||||||||||||| ||
Db        1378 TGCTGAAGCTCATCAATGAAATGGCGCGCACCACCGATATGCTGTGCCGCGAACTCCCCC 1437

Qy        1540 CGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCG 1599
                || | | || |   ||  | ||||||||||| || |  ||  |||||||| |||| ||||
Db        1438 AGGCATTTAACGATCTGGCCGTCGATGGCGTCATTGTTGATCAAATGGAACCGGCAGGCG 1497

Qy        1600 GATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCA 1659
                  | || ||||||||| |||   ||||||||| | |  || ||||||||  ||||  |||
Db        1498 CGCTCGTTGCTGAAGCACTGGGACTGCCGTTTATCTCTGTCGCCTGCGCGCTGCCTCTCA 1557

Qy        1660 ATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAG 1719
               |||||||| | |  || || || ||||| ||||| |||   |  |        |||   |
Db        1558 ATCGTGAACCGGATATGCCCCTGGCGGTTATGCCTTTCGAATACGGGACCAGCGACGCGG 1617

Qy        1720 CGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCG 1779
               | |   |||||| |   |||||     | || |||||  |  | ||||||||||| | |
Db       1618 CTCGCGAACGTTATGCCGCCAGTGAAAAAATTTATGACTGGCTAATGCGTCGTCATGACC 1677

Qy       1780 ACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGT 1839
                || ||   | |||||   |      |    ||  |    ||||      | || ||||
Db       1678 GTGTCATTGCCGAACACAGCCACAGAATGGGCTTAGCCCCCCGGCAAAAGCTTCACCAGT 1737

Qy       1840 GCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGC 1899
              |  | ||||| ||||| |||||||||||| | || || |   | |||||||| ||  |
Db       1738 GTTTTTCGCCACTGGCGCAAATCAGCCAGCTTGTTCCTGAACTGGATTTTCCCCGCAAAG 1797

Qy       1900 AACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGC 1959
                 | || || || | ||| ||||| ||||| || ||||   ||    | ||  |  ||
Db       1798 CGTTACCGGCTTGTTTTCATGCCGTCGGGCCTCTGCGCGAAACGCACGCACCGTCAACGT 1857

Qy       1960 TCCATGCGCCCTG---GCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGC 2016
                    | |  |      || |        | ||   | |||  ||||||||||| ||||
Db       1858 CTTCATCCCGTTATTTTACATCCTCAGAAAAACCCCGGATTTTCGCCTCGCTGGGCACGC 1917

Qy       2017 TGCAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGC 2076
              | || || || || |   ||||||||       | | | | || ||     |  |
Db       1918 TTCAGGGACACCGTTATGGGCTGTTTAAAACGATAGTGAAAGCCTGTGAAGAAATTGACG 1977

Qy       2077 TATCGCTGGTGATCGCCCATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGC 2136
                  |||  || | ||||| |||||  |  | |  | |   |||    |  |||||| ||
Db       1978 GTCAGCTCCTGTTAGCCCACTGTGGTCGTCTTACGGACTCTCAGTGTGAAGAGCTGGCGC 2037

Qy       2137 TCGCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGC 2196
                    |     |   ||||  |||||| | |||||||   || || ||    || ||||
Db       2038 GAAGCCGTCATACACAGGTGGTGGATTTTGCCGATCAGTCAGCCGCGCTGTCTCAGGCGC 2097

Qy       2197 AGCTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGC 2256
              |||||   ||||| || | |||  | || |  |  ||||| ||  |  | | | | ||||
Db       2098 AGCTGGCGATCACCCACGGCGGCATGAATACGGTACTGGACGCGATTAATTACCGGACGC 2157

Qy       2257 CGATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGC 2316
              |  |  | ||||| ||| | || ||||||||||||||||| ||| | ||||| |  |  |
Db       2158 CCCTTTTAGCGCTTCCGCTGGCCTTTGATCAGCCCGGCGTCGCGTCACGCATCGTTTATC 2217

Qy       2317 ATGACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGC 2376
              | | | | ||    || || ||||||| |      |||    |||  |  ||  |||
Db       2218 ACGGCATCGGCAAGCGTGCTTCCCGCTTTACCACCAGCCATGCTTTGGCTCGTCAGATGC 2277

Qy       2377 AACAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGC 2436
                    |||||||| |||     |    ||  || ||| |    || ||| |    || |
Db       2278 GTTCATTGCTGACCAACGTCGACTTTCAGCAGCGCATGGCGAAAATCCAGACAGCCCTTC 2337

Qy       2437 AGCGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGC 2496
                   ||||| ||   |  |   || ||||| ||| | |||||||   |||| || | |
Db       2338 GTTTGGCAGGGGGCACCATGGCCGCTGCCGATATCATTGAGCAGGTTATGTG-CACCGGT 2396

Qy       2497 AAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGC 2556
               || |   ||   |  || |     ||||   |||| || ||| ||||| | || || ||
Db       2397 CAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCGTGGGGGC 2456

Qy       2557 TGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATG 2616
              |||||| ||||||||| | || || |||||||| |  ||   ||| ||  |  ||
Db       2457 TGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATATGCGTAT 2516

Qy       2617 CCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAG 2676
                || |  | ||   ||   | ||  |||  || |||||||| ||||| ||||| |||
Db       2517 TTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTCACCACGA 2576

Qy       2677 CGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGG 2736
               ||| | |  |     |||| ||| ||| |    |||||| |       |  ||| | |
Db       2577 TGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACTGGCCCGA 2636

Qy       2737 TTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGC 2796
               |||||||| || |||||  | |  || || || ||| || | |  || | ||  ||  |
Db       2637 CTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTTGTATTAC 2696

Qy       2797 ATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACAC 2856
               ||      ||| ||      | |  |      | | ||     |  ||||||   | ||
Db       2697 TTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGATGGATAC 2756
```

```
Qy       2857 AGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGC 2916
               || ||   | |||| ||  |       ||   | ||    | || |  |   | | |||
Db       2757 CGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTATCGGTGC 2816

Qy       2917 GCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCA 2976
               |  | ||||||||| || || ||      | | |   ||   |||    |||| |  ||
Db       2817 CCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGGGCTTCCA 2876

Qy       2977 GGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCT 3036
              || |||| |||| || || ||||   ||   || ||||| ||  |   |   || ||  |
Db       2877 GGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTCCCATTAT 2936

Qy       3037 GATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAG 3096
               ||||||||||| ||||||||||||   |||||||| || || ||||  |||||||||
Db       2937 CATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGCCGCTCTC 2996

Qy       3097 CGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAA 3156
                |     |  | || |||||||| || || ||  || |  |  |  |  | |      |
Db       2997 GCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAGATCCTGA 3056

Qy       3157 CACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCT 3216
                 ||| || ||  | ||   ||||||||||   || || || |||  |||    || ||
Db       3057 ATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTCAGACACT 3116

Qy       3217 GCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTG 3276
              |||||| ||||| || |||   ||||| ||||| |||   |||||   |||   ||||||
Db       3117 GCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACGCATTCTG 3176

Qy       3277 GCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCAC 3336
              |   |||||||||  ||  || || || ||  | || ||||| ||||| ||| | |||||
Db       3177 G---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATCCTACCAC 3233

Qy       3337 CGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCAC 3396
              ||| || ||  |||||   || || ||  | || ||     |    || ||      |
Db       3234 CGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTGATGTCTT 3293

Qy       3397 CGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACA 3456
                   |   | |  |   |||    ||    |  ||||||||  |||  ||||   | ||
Db       3294 TACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGCAGCAGCA 3353

Qy       3457 GCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTG 3516
              | | ||||||||  |||| || |||||||||||||| |||||     |  |    |||||
Db       3354 GGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATTCACGCTG 3413

Qy       3517 GCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGG 3576
              ||| || ||||| ||||||||  |  |   ||  |  ||||||  ||| ||||| || ||
Db       3414 GCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTTATGCGGG 3473

Qy       3577 GCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGG 3636
                ||||    |||  ||||      || |||||| ||||||||||||| || ||  |
Db       3474 AAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTCCGGTATT 3533

Qy       3637 TGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACT 3696
               | |||  |||  ||  |  |||   ||        | |   ||  | |||||||  |||
Db       3534 AGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAAACCAACT 3593

Qy       3697 TATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGC 3756
                 || ||||| |||||||| || |||||||| ||||| ||||| || ||||| |||||
Db       3594 ACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGCTGCGGGG 3653

Qy       3757 ATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAG 3816
              || ||   |    | || || || || || || ||||| ||||| || || ||||| |  |||
Db       3654 ATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGTCTACGAG 3713

Qy       3817 GACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGAA 3876
              ||    || |||||||| ||||| || || ||||| |||||||||||||| ||||| |||
Db       3714 GATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGCCATTGAA 3773

Qy       3877 GAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACG 3936
              ||  ||||  | ||||||||||||  | | |  || || ||||| ||| ||||||| |||
Db       3774 GAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCCGGTTACG 3833
```

```
Qy       3937 CCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCCG 3996
Db       3834 CCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGATCAAACC 3893
Qy       3997 CTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTT 4056
Db       3894 CGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCGTCAGTTT 3953
Qy       4057 CTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTT 4116
Db       3954 CTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGTCCCTTTT 4013
Qy       4117 CTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGC 4176
Db       4014 TTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGCATGGAGA 4073
Qy       4177 AGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCC 4236
Db       4074 AGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGCGTTTTCT 4133
Qy       4237 TTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAATT 4296
Db       4134 TTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATACGTTGATA 4193
Qy       4297 CATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTG 4356
Db       4194 CACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGCATTAGTT 4253
Qy       4357 CAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTG 4416
Db       4254 CAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGCCAGAGTC 4313
Qy       4417 AGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGC 4476
Db       4314 AGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGGTCGCAGG 4373
Qy       4477 TTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCTT 4536
Db       4374 TTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGACCTGTTA 4433
Qy       4537 CGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAAC 4596
Db       4434 AGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCATGAGTAAC 4493
Qy       4597 TCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACACC 4656
Db       4494 TCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCATCACACG 4553
Qy       4657 GTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTG 4716
Db       4554 GTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGATGGCCTC 4613
Qy       4717 GCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCG 4776
Db       4614 GCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACTGGCGCCT 4673
Qy       4777 CCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATC 4836
Db       4674 GAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGCGAACCTC 4733
Qy       4837 GACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCAC 4896
Db       4734 GACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGAGCAGCAT 4793
Qy       4897 TACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTTT 4956
Db       4794 TACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTTTGATTTT 4853
Qy       4957 CGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAA 5016
Db       4854 CGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCTTACCCAG 4913
Qy       5017 AGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGT 5076
```

```
Db       4914 AGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCTGGTCGGC 4973
Qy       5077 GCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCC 5136
Db       4974 GCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGCGACAGCA 5033
Qy       5137 AGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACGC 5196
Db       5034 GGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCATGCGGTCG 5090
Qy       5197 AAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGA 5256
Db       5091 AAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGATGCAAAAA 5150
Qy       5257 CGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGATG 5316
Db       5151 CCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTTATTGACG 5210
Qy       5317 GGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGC 5376
Db       5211 ATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAACGTCTGA 5270
Qy       5377 AGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCGT 5436
Db       5271 TGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAACCGGCGT 5330
Qy       5437 TTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGATC 5496
Db       5331 TTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCGTTTGATC 5390
Qy       5497 ATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACGC 5556
Db       5391 ATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGATGATACGC 5450
Qy       5557 TGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCG 5616
Db       5451 TGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATCATGGGCG 5510
Qy       5617 TGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTA 5676
Db       5511 TGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAGTTGACCA 5570
Qy       5677 ACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCT 5736
Db       5571 ATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCGGCAAGCT 5630
Qy       5737 GGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCGC 5796
Db       5631 GGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGTCAGGCGC 5690
Qy       5797 TCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGAT 5856
Db       5691 TGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCTGCCACAG 5750
Qy       5857 CCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTT 5916
Db       5751 CCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAGCAGGTTT 5810
Qy       5917 ATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGC 5976
Db       5811 ACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAGCGGCAGT 5870
Qy       5977 GCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTT 6036
Db       5871 CAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCCCTTACTT 5930
Qy       6037 CGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTT 6096
Db       5931 CCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC------TCT 5984
Qy       6097 ACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAA 6156
```

```
Db          5985 AGCGCCATGTCTTTCCCGGAGCGTCGCCTGAAGTTTTGACAGGGGCGGCGCATAGAGGAA 6044

Qy          6157 ACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGTA 6216
                  ||||| || || || ||||    || |  || || |||||||||||| ||||||||| ||
Db          6045 GCCAAAAGAAACACAACCTTCTTTGCCCCTGACGGCGTGATGCATACGGTGCGCCATATA 6104

Qy          6217 TAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCAG 6276
                  || || || || ||||| |||| || ||||||||| ||||| ||||||||||||||
Db          6105 CAACCGTTTGAGGTAGCCCTTGCGTGGAATATAGCGGAATGGCCAACGTTGATGCACCAG 6164

Qy          6277 GCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACTG 6336
                  || |||||||||||| || |||||    || |||| |||||| || || ||||||||||
Db          6165 CCCGTCGTGCACCATAAAATAGAGTAATCCATACGCCGTCATACCTGCGCCAATCCACTG 6224

Qy          6337 CAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACAC 6396
                  |||||||||||| |||       || | |||||||||||  ||||  | | | ||||| ||
Db          6225 GAGCGGCCACATTCCTGTACTGCCCAGATAAATCAGCAGGATCGATAATGCAGCAAAAAC 6284

Qy          6397 CACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGATG 6456
                 ||| |||||||||||||| |  ||||||  |||  |  ||||||||||| || || |||||
Db          6285 CACGGCATAAAGATCGTTAACTTCAAACGCACCTTTACGCGGTTCATGATGTGAAAGATG 6344

Qy          6457 CCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCAT 6516
                 ||| ||||| ||||| ||||||||||||||||||| || | ||| || || |  | ||||||
Db          6345 CCATCCCCAACCCCAGCCGTGCATGATGTATTTGTGTGCCAGTGCAGCAATCACTTCCAT 6404

Qy          6517 CACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATTGTTCGTCCATT 6576
                   | | ||||||  | || | ||| ||  | ||||| | ||| | |||  ||  |||  |
Db          6405 GCCAATCACGGTAACGAAAACGATCAGGGCATTCCAAATCCACAACATAATTTCTCCGGT 6464

Qy          6577 TGTGGAAAAGGGAAGTA 6593
                  | |      || || |
Db          6465 AGAGACGTCTGGCAGCA 6481

RESULT 2
US-08-980-832-1/c
; Sequence 1, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
;  APPLICANT: Pasamontes, Luis
;  APPLICANT: Tsygankov, Yuri
;  TITLE OF INVENTION: Improved Fermentative Carotenoid Production
;  FILE REFERENCE: Improved Fermentive Carotenoid
;  CURRENT APPLICATION NUMBER: US/08/980,832B
;  CURRENT FILING DATE: 1997-12-01
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 8625
;   TYPE: DNA
;   ORGANISM: Flavobacterium sp. R1534
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8348)..(8349)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8539)..(8540)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8581)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8590)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8592)
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8602)..(8604)
US-08-980-832-1

  Query Match             14.8%;  Score 1035.4;  DB 3;  Length 8625;
  Best Local Similarity   56.5%;  Pred. No. 2.6e-273;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy          2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
                 | ||  || ||     |  |  |  | ||||    |     || |    | ||| || | 
Db          6984 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 6925

Qy          2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
                  | |  || || || ||| |   ||   |||| |||||     | ||  |       || 
Db          6924 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 6865

Qy          2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
                  |   |     | || |||| ||        |     || | | | | || |||||||| 
Db          6864 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 6805

Qy          2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
                 |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||  
Db          6804 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 6745

Qy          2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                    ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db          6744 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 6685

Qy          2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
                  | || ||||   | ||||   |   | | |  ||  |         ||||| |||||||
Db          6684 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 6634

Qy          2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
                  ||   || ||    ||      | |  |    ||    | ||||||| ||| |||   
Db          6633 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 6574

Qy          2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                     | |  ||       ||||| || |  || ||         ||| || || |||   
Db          6573 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 6514

Qy          2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
                  ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db          6513 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 6454

Qy          3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
                 |  |||||  | ||||| || |||||    |||   |  || || || ||  ||||   |
Db          6453 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 6394

Qy          3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145
                 |||||  ||||  ||   ||  |  |||| || ||||| |  |||    ||     ||  
Db          6393 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 6334

Qy          3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205
                 ||||  ||    | |       |||         |||||||||     ||  |||||||| 
Db          6333 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6274

Qy          3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265
                   |     | ||| ||| |||  |  |  |  |  |                   | |  ||| 
Db          6273 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6214

Qy          3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325
                 |     |   |  | ||   ||  ||    | |      || ||| | ||  |||||||| 
Db          6213 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6154

Qy          3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385
                 ||  ||  |||||| || ||  |||| |  |||| |  | | |||||     | || |||
Db          6153 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6094

Qy          3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445
                 |     |  ||    ||   | | |   ||        ||   |      |  ||    |
Db          6093 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6034

Qy          3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505
                    ||||   | || ||  | || ||||| ||||| ||||||||||    | | |   ||| 
Db          6033 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 5977

Qy          3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565
                  | |  ||||  ||| || ||||| || || ||||| ||         | | ||    |||
Db      5976 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 5917

Qy      3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625
Db      5916 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 5857

Qy      3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685
Db      5856 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 5797

Qy      3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745
Db      5796 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 5737

Qy      3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805
Db      5736 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 5677

Qy      3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865
Db      5676 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 5617

Qy      3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925
Db      5616 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 5557

Qy      3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985
Db      5556 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 5497

Qy      3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045
Db      5496 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 5437

Qy      4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105
Db      5436 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 5377

Qy      4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165
Db      5376 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5317

Qy      4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225
Db      5316 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5257

Qy      4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285
Db      5256 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5197

Qy      4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345
Db      5196 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5137

Qy      4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405
Db      5136 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5077

Qy      4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465
Db      5076 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5017

Qy      4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525
Db      5016 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 4957

Qy      4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
Db      4956 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 4897

Qy      4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645
Db      4896 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 4837

Qy      4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705
Db      4836 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 4777

Qy      4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765
Db      4776 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 4717

Qy      4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825
Db      4716 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 4657

Qy      4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885
Db      4656 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 4597

Qy      4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
Db      4596 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 4537

Qy      4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
Db      4536 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 4477

Qy      5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
Db      4476 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 4417

Qy      5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
Db      4416 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 4357

Qy      5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
Db      4356 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4297

Qy      5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
Db      4296 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4240

Qy      5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
Db      4239 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4180

Qy      5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
Db      4179 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4126

Qy      5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
Db      4125 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4066

Qy      5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
Db      4065 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4006

Qy      5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
Db      4005 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 3946

Qy      5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
Db      3945 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 3886

Qy      5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
Db      3885 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 3826

Qy      5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
Db      3825 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 3766
```

```
Qy        5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
               |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db        3765 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 3715

Qy        5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
               |   |  ||||||   ||  |       ||  |  | |  | ||| || |||||||||
Db        3714 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 3655

Qy        5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
               |||||||||    ||  | |||  | | |||| |||||  |||||| | |||||  | ||
Db        3654 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 3595

Qy        5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
                |   |  | ||||||| ||| ||        | | ||| |  ||     | ||||
Db        3594 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 3535

Qy        5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                   |||    ||||    || | ||  ||  | | ||| |||| |   || |
Db        3534 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 3475

Qy        6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
                |||  | | |||||  ||    |    |||    | |  |  || ||||||   || ||
Db        3474 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 3415

Qy        6086 TCG 6088
                ||
Db        3414 GCG 3412


RESULT 3
US-09-920-923B-1/c
; Sequence 1, Application US/09920923B
; Patent No. 6677134
; GENERAL INFORMATION:
;  APPLICANT: Pasamontes, Luis
;  APPLICANT:  Tsygankov, Yuri
;  TITLE OF INVENTION: Fermentative Carotenoid Production
;  FILE REFERENCE: 15464 US (C38435/125944)
;  CURRENT APPLICATION NUMBER: US/09/920,923B
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 08/980,832
;  PRIOR FILING DATE: 1997-12-01
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 8625
;   TYPE: DNA
;   ORGANISM: Flavobacterium sp. R1534
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8348)..(8349)
;   OTHER INFORMATION: unsure
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8539)..(8540)
;   OTHER INFORMATION: unsure
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8581)..(8581)
;   OTHER INFORMATION: unsure
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8590)..(8590)
;   OTHER INFORMATION: unsure
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8592)..(8592)
;   OTHER INFORMATION: unsure
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (8602)..(8604)
;   OTHER INFORMATION: unsure
US-09-920-923B-1
```

```
  Query Match             14.8%;  Score 1035.4;  DB 4;  Length 8625;
  Best Local Similarity   56.5%;  Pred. No. 2.6e-273;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy        2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
               | ||  || ||     |  |  |  | ||||    |     || |    | ||| || |
Db        6984 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 6925

Qy        2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
                | |  || || || ||| |   ||   |||| |||||     | ||  |       ||
Db        6924 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 6865

Qy        2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
                |   |     | || |||| ||        |    ||  | | | | || ||||||||
Db        6864 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 6805

Qy        2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
               |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||
Db        6804 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 6745

Qy        2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                  ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db        6744 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 6685

Qy        2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
                | || ||||   | ||||   |   | | |  ||  |         ||||| |||||||
Db        6684 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 6634

Qy        2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
                ||   || ||    ||      | |  |    ||    | ||||||| |||| |||
Db        6633 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 6574

Qy        2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                   | |  ||       ||||| || |  || ||         ||| || || |||
Db        6573 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 6514

Qy        2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
                ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db        6513 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 6454

Qy        3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
               |  |||||  | ||||| || |||||    |||   |  || || || ||  ||||   |
Db        6453 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 6394

Qy        3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145
               |||||  ||||  ||   ||  |  |||| || ||||| |  |||    ||     ||
Db        6393 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 6334

Qy        3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205
               ||||  ||   | |       |||         |||||||||     ||  ||||||||
Db        6333 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6274

Qy        3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265
                 |     | ||| ||| |||  |  |  | |                   | |  |||
Db        6273 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6214

Qy        3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325
               |     |   |  | ||   ||  ||    | |      || ||| | || ||||||||
Db        6213 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6154

Qy        3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385
               ||  ||  |||||| || ||  |||| |  |||| |  | | |||||     | || |||
Db        6153 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6094

Qy        3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445
               |     |  ||    ||   | | |   ||        ||   |      |  ||    |
Db        6093 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6034

Qy        3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505
                  ||||   | || ||  | || ||||| ||||| ||||||||    | | |   |||
Db        6033 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 5977
```

```
Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565
               | |  ||||  ||| || |||| || || ||||| ||         | | ||    |||
Db       5976 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 5917

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625
              || || |||||||  |||    ||  |||       || ||  |   |||    |||||
Db       5916 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 5857

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685
               | ||  |     | ||  |||||   |||    |      |   |    | ||| |  |
Db       5856 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 5797

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745
              |||    | |    || || |||||||| || ||||| || ||||| || || |||||||
Db       5796 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 5737

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805
              || |||| ||||  | |||||| |  |||||   |||||||| || |||||| |||||||
Db       5736 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 5677

Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865
              |||| |   | ||    |||     ||||||||| || || |||||  | ||||| |||
Db       5676 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 5617

Qy       3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925
               |  | |   |||| |||   | ||    ||  || || |      | ||||   ||||
Db       5616 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 5557

Qy       3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985
              |||||||  ||||||||| || |||   ||||  || |||       | || ||||  |
Db       5556 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 5497

Qy       3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045
              |  |  |      ||||   |  ||| |||||| | ||||||||    ||||| || ||||
Db       5496 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 5437

Qy       4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105
              |||| |  || |  |  ||  |    || || | | | ||||| |||||||||| || || |
Db       5436 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 5377

Qy       4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165
              |   ||| || ||| || || |  | ||||||  || ||||||  | ||   || ||||
Db       5376 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5317

Qy       4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225
              | ||||  |||   ||| ||||||||||||||   || || |||||||   |||||||| |
Db       5316 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5257

Qy       4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285
              |||| || || |||||| |  ||||||| ||||| || || ||| | ||    |||||||
Db       5256 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5197

Qy       4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345
              || |  | || ||||||||||| ||     | ||||| |||||  |    |||||||||
Db       5196 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5137

Qy       4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405
               |  ||||||   ||||||||     |||||||||    | ||||||  |||| | ||||
Db       5136 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5077

Qy       4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465
              |||||   || |  | | | || |||   ||| |||||   |  |  |||||   | |   ||
Db       5076 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5017

Qy       4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525
               ||| || |  ||    || |   | ||||||  ||| | ||||||| |||| | |||
Db       5016 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 4957

Qy       4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
               | | |||||  ||||    |  |  |     |     ||  |    |||||| ||  |||
Db       4956 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 4897

Qy       4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645
              ||  |   |  ||||||||| || ||  ||||| ||||||    |||  |  ||  | || |
Db       4896 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 4837

Qy       4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705
              | ||||||| ||||   || |||||||| ||    | | || |  | ||||| ||||||
Db       4836 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 4777

Qy       4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765
              |    |  ||| | |||||||| ||  | || |||||  ||||||||   | ||||||
Db       4776 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 4717

Qy       4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825
                ||||| ||    ||      ||    ||| ||  | ||||| || ||||||||| |||
Db       4716 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 4657

Qy       4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885
               ||| ||  |||| ||| ||   || |  ||| |||    ||| || || || |   | |
Db       4656 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 4597

Qy       4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
              ||||||  | |  ||| || |   | ||  ||||  |         | | || || |  |
Db       4596 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 4537

Qy       4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
              |    |||||  ||  |   |||   ||||||||||||   || || ||| | |||||||
Db       4536 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 4477

Qy       5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
              |  ||||||||  |||||||||||||||||||||||||||   ||  |||   ||  |||
Db       4476 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 4417

Qy       5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
              |  | ||||| || || ||||||| || |||||| | || || ||  | ||    ||||
Db       4416 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 4357

Qy       5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
              ||||||| ||   | | ||||||  ||| |  ||||    |      | |
Db       4356 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4297

Qy       5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
                  | | ||    |||| ||||  ||| ||| | || |||||    || |||||||| |
Db       4296 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4240

Qy       5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
                   ||  |||      ||| | ||||| ||||||  |  | ||    |||||
Db       4239 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4180

Qy       5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
               ||||| || |||||   | ||||    | |  |  |      || | ||  ||| | ||
Db       4179 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4126

Qy       5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
              ||| ||  ||  |   ||||     || || |    |||      ||  |    |  |
Db       4125 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4066

Qy       5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
                  ||| | ||   |||| | |   | || ||       ||  |  | ||  || | |
Db       4065 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4006

Qy       5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
              | || | || |   | || || ||||| ||||||||      |     | |||   ||| |
Db       4005 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 3946

Qy       5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
               ||||||  ||||   || | ||| |||||||||||| || || ||  ||||||||||||||
Db       3945 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 3886

Qy       5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
               || ||||||||||  |||| |||||||| |||| |||||||||| | || || |||||
Db       3885 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 3826

Qy       5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
              |||||| || ||||| || |||||| |  | || |||||||  |  || ||||||||||||
Db       3825 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 3766

Qy       5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
              |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db       3765 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 3715

Qy       5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
              |   |  ||||||   ||  |       ||  |  | |  | ||| || |||||||||
Db       3714 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 3655

Qy       5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
              |||||||||    ||  | |||  | | |||| |||||  |||||| | |||||  | ||
Db       3654 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 3595

Qy       5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
               |   |  | ||||||| ||| ||        | | ||| |  ||     | ||||
Db       3594 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 3535

Qy       5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                  |||    ||||    || | ||  ||  | | ||| |||| |   || |
Db       3534 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 3475

Qy       6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
               |||  | | |||||  ||    |    |||    | |  |  || ||||||   || ||
Db       3474 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 3415

Qy       6086 TCG 6088
               ||
Db       3414 GCG 3412


RESULT 4
US-08-980-832-27/c
; Sequence 27, Application US/08980832B
; Patent No. 6291204
; GENERAL INFORMATION:
;  APPLICANT: Pasamontes, Luis
;  APPLICANT: Tsygankov, Yuri
;  TITLE OF INVENTION: Improved Fermentative Carotenoid Production
;  FILE REFERENCE: Improved Fermentive Carotenoid
;  CURRENT APPLICATION NUMBER: US/08/980,832B
;  CURRENT FILING DATE: 1997-12-01
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
;   LENGTH: 11233
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Description of Unknown Organism: Plasmid pZea4
US-08-980-832-27

  Query Match             14.8%;  Score 1035.4;  DB 3;  Length 11233;
  Best Local Similarity   56.5%;  Pred. No. 3.1e-273;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy       2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
              | ||  || ||     |  |  |  | |||    |     || |    | ||| || |
Db       7666 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 7607

Qy       2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
               | |  || || || ||| |   ||   |||| |||||     | ||  |       ||
Db       7606 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 7547

Qy       2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
               |   |     | || |||| ||        |    ||  | | | | || ||||||||
Db       7546 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 7487

Qy       2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
              |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||
Db       7486 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 7427

Qy       2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                 ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db       7426 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 7367

Qy       2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
               | || ||||   | ||||   |   | | |  ||  |         ||||| |||||||
Db       7366 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 7316

Qy       2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
               ||   || ||    ||      | |  |    ||    | ||||||| |||| |||
Db       7315 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 7256

Qy       2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                  | |  ||       ||||| || |  || ||         ||| || || |||
Db       7255 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 7196

Qy       2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
               ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db       7195 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 7136

Qy       3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
              |  |||||  | ||||| || |||||    |||   |  || || || ||  ||||   |
Db       7135 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 7076

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145
              |||||  ||||  ||   ||  |  |||| || ||||| |  |||    |     ||
Db       7075 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 7016

Qy       3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205
              ||||  ||   | |       |||          |||||||||     ||  ||||||||
Db       7015 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6956

Qy       3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265
                |     | ||| ||| |||  |  |  |  | |                   | |  |||
Db       6955 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6896

Qy       3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325
              |     |   |  | ||   ||  ||    | |      || ||| | || ||||||||
Db       6895 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6836

Qy       3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385
              ||  ||  |||||| || ||  |||| |  |||| |  | | |||||     | || |||
Db       6835 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6776

Qy       3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445
              |     |  ||    ||   | | |   ||        ||   |      |  ||    |
Db       6775 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6716

Qy       3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505
                 ||||   | || ||  | || ||||| ||||| ||||||||    | | |   |||
Db       6715 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 6659

Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565
               | |  ||||  ||| || ||||| || || ||||| ||         | | ||    |||
Db       6658 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 6599

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625
              || || |||||||  |||    ||  |||       || ||  |   |||    |||||
Db       6598 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 6539

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685
               | ||  |     | ||  |||||   |||    |      |   |    | ||| |  |
Db       6538 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 6479

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745
              |||    | |    || || |||||||| || ||||| ||  ||||| || || |||||||
Db       6478 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 6419

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805
              || |||| ||||||  | |||||| |  |||||   |||||||||| || |||| |||||||
Db       6418 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 6359

Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865
              |||| |   | ||    |||     ||||||||| || || |||||  |  ||||| |||
Db     6358 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 6299

Qy     3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925
Db     6298 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 6239

Qy     3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985
Db     6238 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 6179

Qy     3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045
Db     6178 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 6119

Qy     4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105
Db     6118 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 6059

Qy     4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165
Db     6058 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5999

Qy     4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225
Db     5998 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5939

Qy     4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285
Db     5938 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5879

Qy     4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345
Db     5878 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5819

Qy     4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405
Db     5818 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5759

Qy     4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465
Db     5758 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5699

Qy     4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525
Db     5698 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 5639

Qy     4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
Db     5638 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 5579

Qy     4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645
Db     5578 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 5519

Qy     4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705
Db     5518 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 5459

Qy     4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765
Db     5458 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 5399

Qy     4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825
Db     5398 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 5339

Qy     4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885
Db     5338 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 5279

Qy     4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
Db     5278 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 5219
```

```
Qy     4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
Db     5218 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 5159

Qy     5006 TTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
Db     5158 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 5099

Qy     5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
Db     5098 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 5039

Qy     5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
Db     5038 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4979

Qy     5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
Db     4978 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4922

Qy     5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
Db     4921 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4862

Qy     5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
Db     4861 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4808

Qy     5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
Db     4807 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4748

Qy     5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
Db     4747 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4688

Qy     5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
Db     4687 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 4628

Qy     5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
Db     4627 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 4568

Qy     5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
Db     4567 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 4508

Qy     5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
Db     4507 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 4448

Qy     5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
Db     4447 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 4397

Qy     5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
Db     4396 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 4337

Qy     5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
Db     4336 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 4277

Qy     5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
Db     4276 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 4217

Qy     5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
Db     4216 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 4157
```

```
Qy       6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085

Db       4156 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 4097

Qy       6086 TCG 6088

Db       4096 GCG 4094


RESULT 5
US-09-920-923B-27/c
; Sequence 27, Application US/09920923B
; Patent No. 6677134
; GENERAL INFORMATION:
;  APPLICANT: Pasamontes, Luis
;  APPLICANT:  Tsygankov, Yuri
;  TITLE OF INVENTION: Fermentative Carotenoid Production
;  FILE REFERENCE: 15464 US (C38435/125944)
;  CURRENT APPLICATION NUMBER: US/09/920,923B
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 08/980,832
;  PRIOR FILING DATE: 1997-12-01
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 11233
;   TYPE: DNA
;   ORGANISM: Unknown
;   FEATURE:
;   OTHER INFORMATION: Plasmid pZea4
US-09-920-923B-27

  Query Match             14.8%;  Score 1035.4;  DB 4;  Length 11233;
  Best Local Similarity   56.5%;  Pred. No. 3.1e-273;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy       2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545

Db       7666 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 7607

Qy       2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605

Db       7606 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 7547

Qy       2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665

Db       7546 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 7487

Qy       2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725

Db       7486 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 7427

Qy       2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785

Db       7426 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 7367

Qy       2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845

Db       7366 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 7316

Qy       2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905

Db       7315 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 7256

Qy       2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965

Db       7255 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 7196

Qy       2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025

Db       7195 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 7136

Qy       3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
```

```
Db       7135 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 7076

Qy       3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145

Db       7075 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 7016

Qy       3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205

Db       7015 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6956

Qy       3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265

Db       6955 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6896

Qy       3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325

Db       6895 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6836

Qy       3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385

Db       6835 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6776

Qy       3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445

Db       6775 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6716

Qy       3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505

Db       6715 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 6659

Qy       3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565

Db       6658 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 6599

Qy       3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625

Db       6598 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 6539

Qy       3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685

Db       6538 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 6479

Qy       3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745

Db       6478 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 6419

Qy       3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805

Db       6418 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 6359

Qy       3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865

Db       6358 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 6299

Qy       3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925

Db       6298 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 6239

Qy       3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985

Db       6238 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 6179

Qy       3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045

Db       6178 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 6119

Qy       4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105

Db       6118 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 6059

Qy       4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGACGTCTGC 4165
```

```
Db        6058 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5999

Qy        4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225

Db        5998 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5939

Qy        4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285

Db        5938 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5879

Qy        4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345

Db        5878 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5819

Qy        4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405

Db        5818 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5759

Qy        4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465

Db        5758 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5699

Qy        4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525

Db        5698 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 5639

Qy        4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585

Db        5638 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 5579

Qy        4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645

Db        5578 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 5519

Qy        4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705

Db        5518 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 5459

Qy        4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765

Db        5458 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 5399

Qy        4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825

Db        5398 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 5339

Qy        4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885

Db        5338 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 5279

Qy        4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945

Db        5278 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 5219

Qy        4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005

Db        5218 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 5159

Qy        5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065

Db        5158 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 5099

Qy        5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125

Db        5098 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 5039

Qy        5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185

Db        5038 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4979

Qy        5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245

Db        4978 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4922

Qy        5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305

Db        4921 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4862

Qy        5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365

Db        4861 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4808

Qy        5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425

Db        4807 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4748

Qy        5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485

Db        4747 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4688

Qy        5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545

Db        4687 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 4628

Qy        5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605

Db        4627 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 4568

Qy        5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665

Db        4567 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 4508

Qy        5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725

Db        4507 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 4448

Qy        5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785

Db        4447 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 4397

Qy        5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845

Db        4396 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 4337

Qy        5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905

Db        4336 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 4277

Qy        5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965

Db        4276 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 4217

Qy        5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025

Db        4216 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 4157

Qy        6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085

Db        4156 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 4097

Qy        6086 TCG 6088

Db        4096 GCG 4094
```

RESULT 6
US-07-783-705A-10
; Sequence 10, Application US/07783705A
; Patent No. 5429939
; GENERAL INFORMATION:
; APPLICANT: Misawa, No. 5429939ihiko
; APPLICANT: Kobayashi, Kazuo
; APPLICANT: Nakamura, Katsumi
; APPLICANT: Yamano, Shigeyuki
; TITLE OF INVENTION: DNA SEQUENCES USEFUL FOR THE
; TITLE OF INVENTION: SYNTHESIS OF CAROTENOIDS

```
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Ladas & Parry
;      STREET:  26 West 61 Street
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10023
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720Kb storage
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  N/A
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/783,705A
;      FILING DATE:  19911023
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  JP 1-103078
;      FILING DATE:  21-APR-1989
;      APPLICATION NUMBER:  JP 2-53225
;      FILING DATE:  05-MAR-1990
;      APPLICATION NUMBER:  US 07/519,011
;      FILING DATE:  19-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schwadron, Janet I.
;      REGISTRATION NUMBER:  33,778
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-708-1935
;      TELEFAX:  212-246-5959
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1479 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  other nucleic acid (plasmid DNA)
US-07-783-705A-10

  Query Match             11.8%;  Score 826.2;  DB 1;  Length 1479;
  Best Local Similarity   72.6%;  Pred. No. 2.5e-216;
  Matches 1068;  Conservative    0;  Mismatches  403;  Indels    0;  Gaps    0;

Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              |||||||  |||   || ||||| |||||||| || |||||||| ||||| ||||| ||
Db          1 ATGAAACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTA 60

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              ||||| ||||| || ||   |    | || || || || || ||||| ||||| || ||
Db         61 CAAGCTGCGGGGATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCT 120

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              ||||| |  |||||    || |||||||| ||||| || || ||||| ||||||||||||
Db        121 TATGTCTACGAGGATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCC 180

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
              || ||||| |||||  ||||  | ||||||||||||  | | |  || || ||||| |||
Db        181 AGTGCCATTGAAGAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTG 240

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
               ||||||| ||||| || || |||||||| |||||    || ||     || |||||||
Db        241 CTGCCGGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGAT 300

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              ||  ||||  | | ||| ||   ||||||      ||| ||||| |  ||||| |||||
Db        301 AACGATCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGT 360

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              |||||||| ||||| | ||||||||| |  || |||| ||| || ||||| || |||||
Db        361 TATCGTCAGTTTCTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGT 420

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              || || || ||| |   | |  | |||||||| |||| ||| || ||  |||     |||
Db        421 ACTGTCCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTG 480

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              || |||||| | ||||| ||||| | ||| || |  |  ||  | || || ||||||||
Db        481 CAGGCATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGC 540

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||||||||| ||||||||  || |||||||||| ||||| || || || ||||| ||
Db        541 CAGGCGTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATT 600

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
              ||||| || || || ||||||||||||||| |||||||| ||||||||||| |||||||||
Db        601 TATACGTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACC 660

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
              |||||  | || ||||| |||      |||||  ||||  |||| |||||  |  |  |
Db        661 GGCGCATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTA 720

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
              || |||  ||| |||||  ||||||| |  || ||    |||   | ||| || ||||||
Db        721 AACGCCAGAGTCAGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAG 780

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
              | ||| ||  | |||    |    || || || || ||||| || ||||| ||||||||
Db        781 GACGGTCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTAT 840

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
                | | ||| |  |||| || ||    ||  | || |   | |     ||| ||  ||||
Db        841 CGCGACCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAG 900

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
              |||||||| ||||| |||||||| |||||||||||  ||||||| |  ||||| ||||||
Db        901 CGCATGAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTC 960

Qy       4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
              ||||| ||||| || ||||| ||||||||||| || ||| |||| || || ||||| ||
Db        961 GCGCATCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAAT 1020

Qy       4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
                       || ||||| || || |||||||| ||||||||||||||   ||  ||| ||
Db       1021 CATGATGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCG 1080

Qy       4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
              || ||||| ||    || ||||||||| |  |||||||| ||||||||||||||| | |||
Db       1081 TCACTGGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGC 1140

Qy       4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
              |||||  || ||||||||      || || ||     | ||||| || ||||||||| ||
Db       1141 ACCGCGAACCTCGACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTAC 1200

Qy       4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
              || |||||||| |||||||| || |||||    ||  | || || ||| | |||||||||
Db       1201 CTTGAGCAGCATTACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACG 1260

Qy       4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
              ||||||||||||||||||  |||  ||||| |||| ||||| || |||||  |||||||
Db       1261 CCGTTTGATTTTCGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCC 1320

Qy       5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
               || | || || ||||||||||| || |||||||||||||||      || |  ||||||
Db       1321 GTTCTTACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTC 1380

Qy       5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
              || ||||| || || || |||||||| ||||| ||| | || ||||| ||||| |||||
Db       1381 TACCTGGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCA 1440

Qy       5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATC 5155
              || || || ||  |  |||||||||||||||
Db       1441 AAAGCGACAGCAGGTTTGATGCTGGAGGATC 1471

RESULT 7
US-08-095-726-7
```

```
; Sequence 7, Application US/08095726
; Patent No. 5530188
;  GENERAL INFORMATION:
;    APPLICANT:  Ausich, Rodney L
;    APPLICANT:  Brinkhaus, Friedhelm L
;    APPLICANT:  Mukharji, Indrani
;    APPLICANT:  Proffitt, John H
;    APPLICANT:  Yarger, James G
;    APPLICANT:  Yen, Huei-Che B
;    TITLE OF INVENTION:  Beta-Carotene Biosynthesis in
;    TITLE OF INVENTION:  Genetically Engineered Hosts
;    NUMBER OF SEQUENCES:  79
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amoco Corp., Patents and Licensing Dept
;      STREET:  200 E Randolph St
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60680-0703
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/095,726
;      FILING DATE:  21-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/785,566
;      FILING DATE:  30-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Galloway, No. 5530188val B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3128567180
;      TELEFAX:  3128564972
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1518 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-095-726-7

  Query Match             10.2%;  Score 713.4;  DB 1;  Length 1518;
  Best Local Similarity   67.7%;  Pred. No. 2.4e-185;
  Matches 1034;  Conservative    0;  Mismatches  481;  Indels   12;  Gaps    2;

Qy      3680 AAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTC 3739
             |||  ||| ||   ||   ||||||||||||||||||||| |||||||||||||||||||
Db         2 AAACCATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTC 61

Qy      3740 GCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGAC 3799
             ||||||| ||||| || || ||||||    | || |||||||| ||||| || ||||| |
Db        62 GCCTGCAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTC 121

Qy      3800 GCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCG 3859
             | ||||| || |   | |||   ||||||||||| || ||||| || |||||||||||||
Db       122 GGGCCTACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCG 181

Qy      3860 ATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCG 3919
             |||| | |||  | || | | ||||||| ||||| || |      |    |||||||||
Db       182 ATCCTACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCA 241

Qy      3920 AGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATT 3979
              |||| |||||||||  |||||||| || || ||||||||    || ||    || || |
Db       242 GGCTGCTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACT 301

Qy      3980 ACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAG 4039
             | |  ||  |         ||| |||  ||||||  ||  |||||| || |  || || |
Db       302 ATGCTAACGACAGCTTCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCG 361

Qy      4040 AAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAAC 4099
             | ||||| || |  ||||| || || || |  |  |||||    ||||| ||| ||   |
Db       362 AGGGCTACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCC 421

Qy      4100 TCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGAC 4159
             ||||||  |||||||| ||     | || ||||||||||||| || ||||||| ||
Db       422 TCGGCAGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTA 481

Qy      4160 GTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATC 4219
               || || || ||||  |||||||||     |||  ||   |||||| |||| || ||||
Db       482 AGCTCCAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATC 541

Qy      4220 TGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCAT 4279
             |||| ||||| || || ||||||||  ||||||| ||||| || || ||  | || || |
Db       542 TGCGGCAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGT 601

Qy      4280 CGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCG 4339
             | ||||| ||| | || || || || ||||| || ||||| || ||||| ||    ||||
Db       602 CCATCTACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCG 661

Qy      4340 GCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGT 4399
             ||||||| ||||||||| | ||||||| |   |||||    ||  ||||||| ||| |
Db       662 GCACCGGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCG 721

Qy      4400 TACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGT 4459
              ||| || |||   ||     ||||||       ||  || ||| | |||   || | |
Db       722 AACTCAACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGC 781

Qy      4460 TAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATA 4519
             | | ||  || ||   ||| ||  ||| ||| || ||||| || || ||||||||| | |
Db       782 TCGCGGATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACA 841

Qy      4520 CCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGC 4579
             ||||  | || |||||  ||  |   ||| |||    ||| || ||| ||  |||| |||
Db       842 CCTATAAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAAGCGGGCGGCACGGCTGGAGC 901

Qy      4580 GTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAAC 4639
             | ||| |||||||||||||||||||||| ||||| || |||||||| ||||| |||   |
Db       902 GCAAGAGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCC 961

Qy      4640 AGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTT 4699
             |||| |||||||| ||| |||||||||| || || || || ||| ||||||| ||||| |
Db       962 AGCTGGCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCT 1021

Qy      4700 TCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCG 4759
             | | | |||||  |||||| || || || || || ||||||||| ||||||||   | ||
Db      1022 TTACCGGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCG 1081

Qy      4760 ATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATC 4819
             |||| ||||| || || |||   |||| |||||| || ||| | || |||||||||||||
Db      1082 ATCCCTCGCTCGCGCCTCCCCCGTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATC 1141

Qy      4820 TCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTG 4879
             | |||| |||     | ||||||   ||||| || |||    ||||||| || || ||||
Db      1142 TTGGCAACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTG 1201

Qy      4880 CTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGT 4939
               || || ||  ||| ||| ||||| ||  | |||   ||  | ||||| ||  | || |
Db      1202 ACTACCTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCT 1261

Qy      4940 TTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGG 4999
             |||| | |     ||  ||  || |   |   ||       || |||   ||      |
Db      1262 TTACCCGGCAGACTTCACGACACGCTTGGATCGCGATCTTGGGATCGCTTTTCATCGAGC 1321

Qy      5000 AGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCA 5059
              |||    ||||| ||| ||   |    |         |  ||  || |          |
Db      1322 CGCCTTCGTTGACCCAAGGCTTGTTCGCCGCAAACGCGACACGACATTC---------AA 1372

Qy      5060 ATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTT 5119
             | ||||| |||||||  || ||||| || || ||||||||| | || |||||  | ||
Db      1373 ACCTCTACCTGGTGGCCGCAGGTACTCACCCTGGCGCGGGCATTCCTGGCGTAGTGGGCC 1432

Qy      5120 CGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTT 5179
```

```
                ||| |   ||||||||| |||||| | |||||||      |||||  | || |||  |
Db       1433 TCGCCGAAAGCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCT 1489

Qy       5180 ACTTGAGCAAGTAACGCAAACCATGGC 5206
               ||||| || |  ||||| ||||||||
Db       1490 GCTTGACCACGCCACGCAGACCATGGC 1516


RESULT 8
US-08-096-043-7
; Sequence 7, Application US/08096043
; Patent No. 5530189
;  GENERAL INFORMATION:
;    APPLICANT:  Ausich, Rodney L
;    APPLICANT:  Brinkhaus, Friedhelm L
;    APPLICANT:  Mukharji, Indrani
;    APPLICANT:  Proffitt, John H
;    APPLICANT:  Yarger, James G
;    APPLICANT:  Yen, Huei-Che B
;    TITLE OF INVENTION:  Lycopene Biosynthesis in
;    TITLE OF INVENTION:  Genetically Engineered Hosts
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amoco Corp., Patents and Licensing Dept
;      STREET:  200 E Randolph St
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60680-0703
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,043
;      FILING DATE:  22-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/785,568
;      FILING DATE:  30-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Galloway, No. 5530189val B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3128567180
;      TELEFAX:  3128564972
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1518 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-096-043-7

  Query Match             10.2%;  Score 713.4;  DB 1;  Length 1518;
  Best Local Similarity   67.7%;  Pred. No. 2.4e-185;
  Matches 1034;  Conservative    0;  Mismatches  481;  Indels   12;  Gaps    2;

Qy       3680 AAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTC 3739
              |||  ||| ||   ||   ||||||||||||||||||||| ||||||||||||||||||||
Db          2 AAACCATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTC 61

Qy       3740 GCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGAC 3799
              ||||||| ||||| || || ||||||    | || |||||||| ||||| || ||||| |
Db         62 GCCTGCAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTC 121

Qy       3800 GCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCG 3859
              | ||||| || |   | |||   ||||||||||| || ||||| || ||||||||||||||
Db        122 GGGCCTACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCG 181

Qy       3860 ATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCG 3919
              |||| | |||  | || | | ||||||| ||||| || |      |    |||||||||
```

```
Db        182 ATCCTACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCA 241

Qy       3920 AGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATT 3979
               |||| |||||||||  |||||||| || || ||||||||    || ||    || || |
Db        242 GGCTGCTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACT 301

Qy       3980 ACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAG 4039
              | |  ||  |         ||| |||  ||||||  ||  |||||| || |  || || |
Db        302 ATGCTAACGACAGCTTCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCG 361

Qy       4040 AAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAAC 4099
              | ||||| || |  ||||| || || || |  |  |||||    ||||| ||| ||   |
Db        362 AGGGCTACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCC 421

Qy       4100 TCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGAC 4159
              ||||||  |||||||| ||     | || ||||||||||||| || ||||||| ||
Db        422 TCGGCAGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTA 481

Qy       4160 GTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATC 4219
                || || || ||||  |||||||||     |||  ||    |||||| |||| || ||||
Db        482 AGCTCCAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATC 541

Qy       4220 TGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCAT 4279
              |||| ||||| || || ||||||||  ||||||| ||||| || || ||  | || || |
Db        542 TGCGGCAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGT 601

Qy       4280 CGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCG 4339
              | ||||| ||| | || || || || ||||| || ||||| || ||||| ||    ||||
Db        602 CCATCTACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCG 661

Qy       4340 GCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGT 4399
              ||||||| |||||||||| | ||||||| |   |||||     ||  ||||||| ||| |
Db        662 GCACCGGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCG 721

Qy       4400 TACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGT 4459
               ||| || ||   ||      ||||||       ||  || ||| | |||   || | |
Db        722 AACTCAACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGC 781

Qy       4460 TAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATA 4519
              | | ||  || ||   ||| ||  ||| ||| || ||||| || || ||||||||| | |
Db        782 TCGCGGATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACA 841

Qy       4520 CCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGC 4579
              ||||  | || |||||  ||  |   ||| |||    ||| || ||| ||  |||| |||
Db        842 CCTATAAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAAGCGGGCGGCACGGCTGGAGC 901

Qy       4580 GTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAAC 4639
              | ||| |||||||||||||||||||||| ||||| || |||||||| |||||| ||| |||   |
Db        902 GCAAGAGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCC 961

Qy       4640 AGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTT 4699
              |||| |||||||| ||| |||||||||| || || || ||  ||| |||||||| ||||| |
Db        962 AGCTGGCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCT 1021

Qy       4700 TCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCG 4759
              | | | |||||  |||||| || || || || || |||||||||| ||||||||   | ||
Db       1022 TTACCGGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCG 1081

Qy       4760 ATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATC 4819
              |||| ||||| || || |||   |||| |||||| || ||| | || |||||||||||||
Db       1082 ATCCCTCGCTCGCGCCTCCCCCGTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATC 1141

Qy       4820 TCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTG 4879
              | |||| |||     | ||||||   ||||| || |||     ||||||| || || ||||
Db       1142 TTGGCAACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTG 1201

Qy       4880 CTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGT 4939
                || || ||  ||| ||| ||||| ||  | |||   ||  | ||||| ||  | || |
Db       1202 ACTACCTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCT 1261

Qy       4940 TTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGG 4999
              |||| | |     ||  ||  || |   |   ||        || |||   ||      |
Db       1262 TTACCCGGCAGACTTCACGACACGCTTGGATCGCGATCTTGGGATCGCTTTTCATCGAGC 1321
```

```
Qy    5000 AGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCA 5059
            |||    ||||| ||| ||   |    |         |  ||  || |          |
Db    1322 CGCCTTCGTTGACCCAAGGCTTGTTCGCCGCAAACGCGACACGACATTC---------AA 1372

Qy    5060 ATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTT 5119
           | ||||| |||||||  || ||||| || || ||||||||| | || |||||  | ||  
Db    1373 ACCTCTACCTGGTGGCCGCAGGTACTCACCCTGGCGCGGGCATTCCTGGCGTAGTGGGCC 1432

Qy    5120 CGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTT 5179
             ||| |   ||||||||| |||||| | |||||||      |||||  | || |||  |
Db    1433 TCGCCGAAAGCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCT 1489

Qy    5180 ACTTGAGCAAGTAACGCAAACCATGGC 5206
            ||||| || |  ||||| ||||||||
Db    1490 GCTTGACCACGCCACGCAGACCATGGC 1516

RESULT 9
US-08-096-623A-7
; Sequence 7, Application US/08096623A
; Patent No. 5684238
;  GENERAL INFORMATION:
;    APPLICANT:  Ausich, Rodney L.
;    APPLICANT:  Brinkhaus, Friedhelm L.
;    APPLICANT:  Mukharji, Indrani
;    APPLICANT:  Proffitt, John H.
;    APPLICANT:  Yarger, James G.
;    APPLICANT:  Yen, Huei-Che B.
;    TITLE OF INVENTION:  Biosynthesis of Zeaxanthin and
;    TITLE OF INVENTION:  Glycosylated Zeaxanthin in Genetically Engineered Hosts
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Welsh & Katz, Ltd.
;      STREET:  120 S. Riverside Plaza, 22nd Floor
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,623A
;      FILING DATE:  22-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/805,061
;      FILING DATE:  09-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/662,921
;      FILING DATE:  28-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/562,674
;      FILING DATE:  03-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/525,551
;      FILING DATE:  18-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/487,613
;      FILING DATE:  02-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gamson, Edward P.
;      REGISTRATION NUMBER:  29,381
;      REFERENCE/DOCKET NUMBER:  AMO-006.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 655-1500
;      TELEFAX:  (312) 655-1501
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1518 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    POSITION IN GENOME:
;      MAP POSITION:  -6 to 1512
;      UNITS:  bp
US-08-096-623A-7

  Query Match             10.2%;  Score 713.4;  DB 1;  Length 1518;
  Best Local Similarity   67.7%;  Pred. No. 2.4e-185;
  Matches 1034;  Conservative    0;  Mismatches  481;  Indels   12;  Gaps    2;

Qy    3680 AAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTC 3739
           |||  ||| ||   ||   ||||||||||||||||||||| |||||||||||||||||||
Db       2 AAACCATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTC 61

Qy    3740 GCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGAC 3799
           ||||||| ||||| || || ||||||    | || |||||||| ||||| || ||||| |
Db      62 GCCTGCAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTC 121

Qy    3800 GCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCG 3859
           | ||||| || |   | |||   ||||||||||| || ||||| || |||||||||||||
Db     122 GGGCCTACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCG 181

Qy    3860 ATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCG 3919
           |||| | |||  | || | | ||||||| ||||| || |      |    ||||||||| 
Db     182 ATCCTACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCA 241

Qy    3920 AGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATT 3979
            |||| |||||||||  |||||||| || || ||||||||    || ||    || || |
Db     242 GGCTGCTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACT 301

Qy    3980 ACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAG 4039
           | |  ||  |         ||| |||  ||||||  ||  |||||| || |  || || |
Db     302 ATGCTAACGACAGCTTCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCG 361

Qy    4040 AAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAAC 4099
           | ||||| || |  ||||| || || || |  |  |||||    ||||| ||| ||   |
Db     362 AGGGCTACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCC 421

Qy    4100 TCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGAC 4159
           ||||||  |||||||| ||     | || ||||||||||||| ||  |||||| ||    
Db     422 TCGGCAGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTA 481

Qy    4160 GTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATC 4219
             || || || ||||  |||||||||     |||  ||   |||||| |||| || ||||
Db     482 AGCTCCAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATC 541

Qy    4220 TGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCAT 4279
           |||| ||||| || || ||||||||  ||||||| ||||| || || ||  | || || |
Db     542 TGCGGCAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGT 601

Qy    4280 CGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCG 4339
           | ||||| ||| | || || || || ||||| || ||||| || ||||| ||    ||||
Db     602 CCATCTACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCG 661

Qy    4340 GCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGT 4399
           ||||||| ||||||||| | ||||||| |   |||||    ||  ||||||| ||| |  
Db     662 GCACCGGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCG 721

Qy    4400 TACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGT 4459
            ||| || |||   ||     ||||||        ||  || ||| | |||   || | | 
Db     722 AACTCAACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGC 781

Qy    4460 TAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATA 4519
           | | ||  || ||   |||  ||  ||| ||| || ||||| || || ||||||||| | |
Db     782 TCGCGGATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACA 841

Qy    4520 CCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGC 4579
           ||||  | || |||||  ||  |   ||| |||    ||| || ||| ||  |||| |||
Db     842 CCTATAAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAAGCGGGCGGCACGGCTGGAGC 901
```

```
Qy       4580 GTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAAC 4639
              | ||| |||||||||||||||||||||| ||||| || |||||||| ||||| |||   |
Db        902 GCAAGAGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCC 961

Qy       4640 AGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTT 4699
              |||| |||||||| ||| |||||||||| || || || || ||| ||||||| ||||| |
Db        962 AGCTGGCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCT 1021

Qy       4700 TCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCG 4759
              | | | |||||  |||||| || || || || || ||||||||| ||||||||   | ||
Db       1022 TTACCGGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCG 1081

Qy       4760 ATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATC 4819
              |||| ||||| || || |||   |||| |||||| || ||| | || |||||||||||||
Db       1082 ATCCCTCGCTCGCGCCTCCCCCGTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATC 1141

Qy       4820 TCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTG 4879
              | |||| |||     | ||||||   ||||| || |||    ||||||| || || ||||
Db       1142 TTGGCAACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTG 1201

Qy       4880 CTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGT 4939
                || || ||  ||| ||| ||||| ||  | |||   ||  | ||||| ||  | || |
Db       1202 ACTACCTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCT 1261

Qy       4940 TTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGG 4999
              |||| | |     ||  ||  || |   |   ||       || |||   ||      | 
Db       1262 TTACCCGGCAGACTTCACGACACGCTTGGATCGCGATCTTGGGATCGCTTTTCATCGAGC 1321

Qy       5000 AGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCA 5059
               |||    ||||| ||| ||   |    |         |  ||  || |          |
Db       1322 CGCCTTCGTTGACCCAAGGCTTGTTCGCCGCAAACGCGACACGACATTC---------AA 1372

Qy       5060 ATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTT 5119
              | ||||| |||||||  || ||||| || || ||||||||| | || |||||  | ||  
Db       1373 ACCTCTACCTGGTGGCCGCAGGTACTCACCCTGGCGCGGGCATTCCTGGCGTAGTGGGCC 1432

Qy       5120 CGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTT 5179
                ||| |   ||||||||| |||||| | |||||||      |||||  | || |||  |
Db       1433 TCGCCGAAAGCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCT 1489

Qy       5180 ACTTGAGCAAGTAACGCAAACCATGGC 5206
               ||||| || |  ||||| ||||||||
Db       1490 GCTTGACCACGCCACGCAGACCATGGC 1516

RESULT 10
US-08-095-726-9
; Sequence 9, Application US/08095726
; Patent No. 5530188
;  GENERAL INFORMATION:
;    APPLICANT:  Ausich, Rodney L
;    APPLICANT:  Brinkhaus, Friedhelm L
;    APPLICANT:  Mukharji, Indrani
;    APPLICANT:  Proffitt, John H
;    APPLICANT:  Yarger, James G
;    APPLICANT:  Yen, Huei-Che B
;    TITLE OF INVENTION:  Beta-Carotene Biosynthesis in
;    TITLE OF INVENTION:  Genetically Engineered Hosts
;    NUMBER OF SEQUENCES:  79
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amoco Corp., Patents and Licensing Dept
;      STREET:  200 E Randolph St
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60680-0703
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/095,726
;      FILING DATE:  21-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/785,566
;      FILING DATE:  30-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Galloway, No. 5530188val B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3128567180
;      TELEFAX:  3128564972
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1522 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-095-726-9

  Query Match             10.1%;  Score 708.4;  DB 1;  Length 1522;
  Best Local Similarity   67.6%;  Pred. No. 5.6e-184;
  Matches 1029;  Conservative    0;  Mismatches  481;  Indels   12;  Gaps    2;

Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              ||| ||   ||   ||||||||||||||||||||| ||||||||||||||||||||||||
Db         11 ATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTCGCCTG 70

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              || ||||| || || ||||||    | || |||||||| ||||| || ||||| || |||
Db         71 CAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTCGGGCC 130

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              || || |   | |||   ||||||||||| || ||||| || ||||||||||||||||| 
Db        131 TACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCGATCCT 190

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
              | |||  | || | | ||||||| ||||| || |      |    |||||||||  ||||
Db        191 ACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCAGGCTG 250

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
               |||||||||  |||||||| || || ||||||||    || ||    || || || |  
Db        251 CTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACTATGCT 310

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              ||  |         ||| ||| ||||||  ||  |||||| || |  || || || |||
Db        311 AACGACAGCTTCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCGAGGGC 370

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              || || |  ||||| || || || |  |  |||||    ||||| ||| ||   ||||||
Db        371 TACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCCTCGGC 430

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              |  |||||||| ||     | || ||||||||||||| || ||||||| ||      || 
Db        431 AGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTAAGCTC 490

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              || || ||||  |||||||||     |||  ||   |||||| |||| || |||||||| 
Db        491 CAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATCTGCGG 550

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||| || || ||||||||  ||||||| ||||| || || ||  | || || || |||
Db        551 CAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGTCCATC 610

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
              || ||| | || || || || ||||| || ||||| || ||||| ||    |||||||||
Db        611 TACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCGGCACC 670

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
              || ||||||||| | ||||||| |   |||||    ||  ||||||| ||| |   ||| 
Db        671 GGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCGAACTC 730

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
              || |||   ||     ||||||       ||  || ||| | |||   || | | | | |
Db        731 AACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGCTCGCG 790

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
              |  || ||   ||| ||  ||| ||| || ||||| || || ||||||||| | |||||
Db        791 GATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACACCTAT 850

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
               | || |||||  ||  |   ||| |||    ||| || ||| ||  |||| |||| |||
Db        851 AAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAAGCGGGCGGCACGGCTGGAGCGCAAG 910

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
               |||||||||||||||||||||| ||||| || |||||||| ||||| |||   |||||
Db        911 AGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCCAGCTG 970

Qy       4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
              |||||||| ||| |||||||||| || || || || ||| ||||||| ||||| || | |
Db        971 GCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCTTTACC 1030

Qy       4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
               |||||  |||||| || || || || || ||||||||| ||||||||   | |||||||
Db       1031 GGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCGATCCC 1090

Qy       4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
              ||||| || || |||   |||| |||||| || ||| | || |||||||||||||| |||
Db       1091 TCGCTCGCGCCTCCCCCGTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATCTTGGC 1150

Qy       4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
              | |||     | ||||||   ||||| || |||    ||||||| || || ||||  ||
Db       1151 AACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTGACTAC 1210

Qy       4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
              || ||  ||| ||| ||||| ||  | |||   ||  | ||||| ||  | || |||||
Db       1211 CTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCTTTACC 1270

Qy       4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
              | |     ||  ||  || |   |   ||       || |||   ||      |  |||
Db       1271 CGGCAGACTTCACGACACGCTTGGATCGCGATCTTGGGATCGCTTTTCATCGAGCCGCCT 1330

Qy       5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
                 ||||| ||| ||   |    |         |  ||  || |          || |||
Db       1331 TCGTTGACCCAAGGCTTGTTCGCCGCAAACGCGACACGACATTC---------AAACCTC 1381

Qy       5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
              || |||||||  || ||||| || || ||||||||| | || |||||  | ||    |||
Db       1382 TACCTGGTGGCCGCAGGTACTCACCCTGGCGCGGGCATTCCTGGCGTAGTGGGCCTCGCC 1441

Qy       5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTG 5184
               |   ||||||||| |||||| | |||||||      |||||  | || |||  | ||||
Db       1442 GAAAGCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCTGCTTG 1498

Qy       5185 AGCAAGTAACGCAAACCATGGC 5206
              | || |  |||   ||||||||
Db       1499 ACCACGCCACGTCGACCATGGC 1520

RESULT 11
US-08-096-043-9
; Sequence 9, Application US/08096043
; Patent No. 5530189
;  GENERAL INFORMATION:
;    APPLICANT:  Ausich, Rodney L
;    APPLICANT:  Brinkhaus, Friedhelm L
;    APPLICANT:  Mukharji, Indrani
;    APPLICANT:  Proffitt, John H
;    APPLICANT:  Yarger, James G
;    APPLICANT:  Yen, Huei-Che B
;    TITLE OF INVENTION:  Lycopene Biosynthesis in
;    TITLE OF INVENTION:  Genetically Engineered Hosts
;    NUMBER OF SEQUENCES:  70
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Amoco Corp., Patents and Licensing Dept
;      STREET:  200 E Randolph St
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60680-0703
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,043
;      FILING DATE:  22-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/785,568
;      FILING DATE:  30-OCT-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Galloway, No. 5530189val B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  3128567180
;      TELEFAX:  3128564972
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1522 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-096-043-9

  Query Match             10.1%;  Score 708.4;  DB 1;  Length 1522;
  Best Local Similarity   67.6%;  Pred. No. 5.6e-184;
  Matches 1029;  Conservative    0;  Mismatches  481;  Indels   12;  Gaps    2;

Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              ||| ||   ||   ||||||||||||||||||||| ||||||||||||||||||||||||
Db         11 ATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTCGCCTG 70

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              || ||||| || || ||||||    | || |||||||| ||||| || ||||| || |||
Db         71 CAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTCGGGCC 130

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              || || |   | |||   ||||||||||| || ||||| || |||||||||||||||||
Db        131 TACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCGATCCT 190

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
              | |||  | || | | ||||||| ||||| || |      |    |||||||||  ||||
Db        191 ACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCAGGCTG 250

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
               |||||||||  |||||||| || || ||||||||    || ||    || || || |
Db        251 CTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACTATGCT 310

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              ||  |         ||| |||  ||||||  ||  |||||| || |  || || || |||
Db        311 AACGACAGCTTCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCGAGGGC 370

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              || || |  ||||| || || || |  |  |||||    ||||| ||| ||   ||||||
Db        371 TACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCCTCGGC 430

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              |  |||||||| ||     | || ||||||||||||| || ||||||| ||      ||
Db        431 AGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTAAGCTC 490

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              || || ||||  |||||||||     |||  ||   |||||| |||| || ||||||||
Db        491 CAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATCTGCGG 550

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||| || || ||||||||  ||||||| ||||| || || ||  | || || || |||
Db        551 CAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGTCCATC 610
```

```
Qy        4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
               || ||| | || || || || ||||| || ||||| || ||||| ||    |||||||||
Db         611 TACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCGGCACC 670

Qy        4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
               || ||||||||| | ||||||| |   |||||    ||  ||||||| ||| |   |||
Db         671 GGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCGAACTC 730

Qy        4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
               || |||   ||     ||||||       ||  || ||| | |||   || | | | | |
Db         731 AACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGCTCGCG 790

Qy        4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
               |  || ||   ||| ||  ||| ||| || ||||| || || ||||||||| | |||||
Db         791 GATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACACCTAT 850

Qy        4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
                | || |||||  ||  |   ||| |||    ||| || ||| ||  |||| |||| |||
Db         851 AAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAAGCGGGCGGCACGGCTGGAGCGCAAG 910

Qy        4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
                |||||||||||||||||||||| ||||| || |||||||| ||||| |||   |||||
Db         911 AGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCCAGCTG 970

Qy        4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
               |||||||| ||| |||||||||| || || || || ||| ||||||| ||||| || | |
Db         971 GCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCTTTACC 1030

Qy        4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
                |||||  |||||| || || || || ||||||||| ||||||||   | ||||||
Db        1031 GGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCGATCCC 1090

Qy        4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
               ||||| || || |||   |||| |||||| || ||| | || |||||||||||||| |||
Db        1091 TCGCTCGCGCCTCCCCCGTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATCTTGGC 1150

Qy        4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
               | |||     | ||||||   ||||| || |||    ||||||| || || ||||  ||
Db        1151 AACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTGACTAC 1210

Qy        4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
               || ||  ||| ||| ||||| ||  | |||   ||  | ||||| ||  | || |||||
Db        1211 CTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCTTTACC 1270

Qy        4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
               | |     ||  ||  || |   |   ||       || |||   ||      |  |||
Db        1271 CGGCAGACTTCACGACACGCTTGGATCGCGATCTTGGGATCGCTTTTCATCGAGCCGCCT 1330

Qy        5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
                  ||||| ||| ||   |    |         |  ||  || |          || |||
Db        1331 TCGTTGACCCAAGGCTTGTTCGCCGCAAACGCGACACGACATTC---------AAACCTC 1381

Qy        5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
               || |||||||  || ||||| || || ||||||||| | || |||||  | ||    |||
Db        1382 TACCTGGTGGCCGCAGGTACTCACCCTGGCGCGGGCATTCCTGGCGTAGTGGGCCTCGCC 1441

Qy        5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTG 5184
                |   ||||||||| |||||| | |||||||      |||||  | || |||  | ||||
Db        1442 GAAAGCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCTGCTTG 1498

Qy        5185 AGCAAGTAACGCAAACCATGGC 5206
               | || |  |||   ||||||||
Db        1499 ACCACGCCACGTCGACCATGGC 1520

RESULT 12
US-08-096-623A-9
; Sequence 9, Application US/08096623A
; Patent No. 5684238
;  GENERAL INFORMATION:
;    APPLICANT:  Ausich, Rodney L.
;    APPLICANT:  Brinkhaus, Friedhelm L.
;    APPLICANT:  Mukharji, Indrani
;    APPLICANT:  Proffitt, John H.
;    APPLICANT:  Yarger, James G.
;    APPLICANT:  Yen, Huei-Che B.
;    TITLE OF INVENTION:  Biosynthesis of Zeaxanthin and
;    TITLE OF INVENTION:  Glycosylated Zeaxanthin in Genetically Engineered Hosts
;    NUMBER OF SEQUENCES:  104
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Welsh & Katz, Ltd.
;      STREET:  120 S. Riverside Plaza, 22nd Floor
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/096,623A
;      FILING DATE:  22-JUL-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/805,061
;      FILING DATE:  09-DEC-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/662,921
;      FILING DATE:  28-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/562,674
;      FILING DATE:  03-AUG-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/525,551
;      FILING DATE:  18-MAY-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/487,613
;      FILING DATE:  02-MAR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gamson, Edward P.
;      REGISTRATION NUMBER:  29,381
;      REFERENCE/DOCKET NUMBER:  AMO-006.1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 655-1500
;      TELEFAX:  (312) 655-1501
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1522 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    POSITION IN GENOME:
;      MAP POSITION:  -10 to 1512
;      UNITS:  bp
US-08-096-623A-9

  Query Match             10.1%;  Score 708.4;  DB 1;  Length 1522;
  Best Local Similarity   67.6%;  Pred. No. 5.6e-184;
  Matches 1029;  Conservative    0;  Mismatches  481;  Indels   12;  Gaps    2;

Qy        3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
               ||| ||   ||   ||||||||||||||||||||| ||||||||||||||||||||||||
Db          11 ATGGAAAAAACCGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTCGCCTG 70

Qy        3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
               || ||||| || || ||||||    | || |||||||| ||||| || ||||| || |||
Db          71 CAGGCGGCAGGGATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTCGGGCC 130

Qy        3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
               || || |   | |||   ||||||||||| || ||||| || |||||||||||||||||
Db         131 TACGTCTGGCATGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCGATCCT 190

Qy        3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
```

```
               | |||  | || | | ||||||| ||||| || |      |    |||||||||  ||||
Db         191 ACCGCGCTTGAGGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCAGGCTG 250

Qy        3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
                |||||||||  |||||||| || || ||||||||    || ||    || || || |
Db         251 CTGCCGGTAAAACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACTATGCT 310

Qy        3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
               ||  |         ||| |||  ||||||  ||  |||||| || |  || || || |||
Db         311 AACGACAGCTTCGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCGAGGGC 370

Qy        4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
               || || |  ||||| || || || |  |  |||||    ||||| ||| ||   ||||||
Db         371 TACCGGCGCTTTCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCCTCGGC 430

Qy        4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
               |  |||||||| ||     | || ||||||||||||| || ||||||| ||      ||
Db         431 AGCGTGCCGTTCCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTAAGCTC 490

Qy        4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
               || || ||||  |||||||||     |||  ||   |||||| |||| || ||||||||
Db         491 CAGGCGTGGCAGAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATCTGCGG 550

Qy        4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
               ||||| || || ||||||||  ||||||| ||||| || || ||  | || || || |||
Db         551 CAGGCCTTCTCGTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGTCCATC 610

Qy        4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
               || ||| | || || || || ||||| || ||||| || ||||| ||    |||||||||
Db         611 TACACCCTGATCCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCGGCACC 670

Qy        4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
               || ||||||||| | ||||||| |   |||||    ||  ||||||| ||| |   |||
Db         671 GGGGCGCTGGTGAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCGAACTC 730

Qy        4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
               || |||   ||     ||||||       ||  || ||| | |||   || | | | | |
Db         731 AACGCCCGGGTCGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGCTCGCG 790

Qy        4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
               |  || ||   ||| ||  ||| ||| || ||||| || || ||||||||| | |||||
Db         791 GATGGTCGGATCTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACACCTAT 850

Qy        4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
                | || |||||  ||  |   ||| |||    ||| || ||| ||  |||| |||| |||
Db         851 AAAAAGCTGCTCGGCACCATACCGGTGGGGCAGAAGCGGGCGGCACGGCTGGAGCGCAAG 910

Qy        4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
                |||||||||||||||||||||| ||||| || |||||||| ||||| |||   |||||
Db         911 AGCATGAGCAACTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCCAGCTG 970

Qy        4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
               |||||||| ||| |||||||||| || || || || ||| ||||||| ||||| || | |
Db         971 GCGCACCATACCATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCTTTACC 1030

Qy        4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
                |||||  |||||| || || || || || ||||||||| ||||||||   | ||||||
Db        1031 GGCAGCGCGCTGGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCGATCCC 1090

Qy        4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
               ||||| || || |||   |||| |||||| || ||| | || |||||||||||||| |||
Db        1091 TCGCTCGCGCCTCCCCCGTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATCTTGGC 1150

Qy        4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
               | |||     | ||||||   ||||| || |||    ||||||| || || ||||  ||
Db        1151 AACGCGCCGCTGGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTGACTAC 1210

Qy        4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
               || ||  ||| ||| ||||| ||  | |||   ||  | ||||| ||  | || |||||
Db        1211 CTTGAAGAGCGCTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCTTTACC 1270

Qy        4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
               | |     ||  ||  || |   |   ||       || |||   ||      |  |||
Db        1271 CGGCAGACTTCACGACACGCTTGGATCGCGATCTTGGGATCGCTTTTCATCGAGCCGCCT 1330

Qy        5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
                  ||||| ||| ||   |    |         |  ||  || |          || |||
Db        1331 TCGTTGACCCAAGGCTTGTTCGCCGCAAACGCGACACGACATTC---------AAACCTC 1381

Qy        5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
               || |||||||  || ||||| || || ||||||||| | || |||||  | ||    |||
Db        1382 TACCTGGTGGCCGCAGGTACTCACCCTGGCGCGGGCATTCCTGGCGTAGTGGGCCTCGCC 1441

Qy        5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTG 5184
                |   ||||||||| |||||| | |||||||      |||||  | || |||  | ||||
Db        1442 GAAAGCACCGCCAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCTGCTTG 1498

Qy        5185 AGCAAGTAACGCAAACCATGGC 5206
               | || |  |||   ||||||||
Db        1499 ACCACGCCACGTCGACCATGGC 1520

RESULT 13
US-08-660-645A-6
; Sequence 6, Application US/08660645A
; Patent No. 6087152
;  GENERAL INFORMATION:
;    APPLICANT:  Hohmann, Hans-Peter
;    APPLICANT:  Pasamontes, Luis
;    APPLICANT:  Tessier, Michel
;    APPLICANT:  van Loon, Adolphus
;    TITLE OF INVENTION:  FERMENTATIVE CAROTENOID PRODUCTION
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/660,645A
;      FILING DATE:  07-JUN-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 95108888.9
;      FILING DATE:  09-JUN-1995
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pokras, Bruce A.
;      REGISTRATION NUMBER:  32,748
;      REFERENCE/DOCKET NUMBER:  RAN 6002/170
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 235-5801
;      TELEFAX:  (201) 235-2363
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-660-645A-6

  Query Match             8.8%;  Score 619.4;  DB 3;  Length 1482;
  Best Local Similarity   63.7%;  Pred. No. 1.5e-159;
  Matches  941;  Conservative    0;  Mismatches  536;  Indels    0;  Gaps    0;

Qy        3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
               ||||    | |    || || |||||||| || ||||| || ||||| || || ||||||
Db           1 ATGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTG 60
```

```
Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
Db         61 CAATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCC 120

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
Db        121 TATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCC 180

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
Db        181 GACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTG 240

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
Db        241 CTGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTG 300

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
Db        301 AACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGC 360

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
Db        361 TATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGG 420

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
Db        421 ACCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTG 480

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
Db        481 CAGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGG 540

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
Db        541 CAGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATC 600

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
Db        601 TATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACC 660

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
Db        661 AACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTG 720

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
Db        721 AATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTG 780

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
Db        781 GACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTAT 840

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
Db        841 CGCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAG 900

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
Db        901 CGCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTG 960

Qy       4645 GCGCACCACACCGTCTGTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
Db        961 GCCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAAC 1020

Qy       4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
Db       1021 GGGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCC 1080

Qy       4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
Db       1081 AGCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGC 1140

Qy       4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
Db       1141 CGCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAA 1200

Qy       4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
Db       1201 CTGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGC 1260

Qy       4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
Db       1261 CCCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCG 1320

Qy       5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
Db       1321 ATCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTC 1380

Qy       5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
Db       1381 TACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCC 1440

Qy       5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCG 5161
Db       1441 AAGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCG 1477
```

```
RESULT 14
US-09-298-718-6
; Sequence 6, Application US/09298718
; Patent No. 6124113
;  GENERAL INFORMATION:
;    APPLICANT:  Hohmann, Hans-Peter
;    APPLICANT:  Pasamontes, Luis
;    APPLICANT:  Tessier, Michel
;    APPLICANT:  van Loon, Adolphus
;    TITLE OF INVENTION:  FERMENTATIVE CAROTENOID PRODUCTION
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/298,718
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/660,645
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pokras, Bruce A.
;      REGISTRATION NUMBER:  32,748
;      REFERENCE/DOCKET NUMBER:  RAN 6002/170
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 235-5801
;      TELEFAX:  (201) 235-2363
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-09-298-718-6

  Query Match             8.8%;  Score 619.4;  DB 3;  Length 1482;
  Best Local Similarity   63.7%;  Pred. No. 1.5e-159;
  Matches  941;  Conservative    0;  Mismatches  536;  Indels    0;  Gaps    0;
```

```
Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              ||||    | |    || || |||||||| || ||||| || ||||| || || ||||||
Db          1 ATGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTG 60

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              ||| |||| |||||  | |||||| |  |||||   |||||||| || ||||| ||||||
Db         61 CAATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCC 120

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              ||||| |   | ||    |||     ||||||||| || || |||||  | ||||| |||
Db        121 TATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCC 180

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
                |  | |   |||| |||   | ||    ||  || || |      | ||||   ||||
Db        181 GACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTG 240

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
               |||||||  |||||||||| || |||   ||||  || |||       | || ||||
Db        241 CTGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTG 300

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              ||  |  |      ||||   |  ||| ||||| | ||||||||    ||||| || |||
Db        301 AACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGC 360

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              ||||| |  || |  |  ||  |    || || | | | ||||| |||||||| || ||
Db        361 TATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGG 420

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              ||   ||| || ||| || || |  | ||||||  || |||||||  | ||   || |||
Db        421 ACCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTG 480

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              || ||||  |||   ||| ||||||||||||||   || || ||||||   ||||||||
Db        481 CAGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGG 540

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||| || || |||||| |  ||||||| ||||| || || ||| | ||    ||||||
Db        541 CAGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATC 600

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
              ||| |  | || ||||||||||| ||     | ||||| |||||  |    |||||||||
Db        601 TATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACC 660

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
                |  ||||||   ||||||||     |||||||||    | ||||||  |||| | |||
Db        661 AACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTG 720

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
              ||||||   || |  | | | || |||   ||| |||||   |  |||||   | |   |
Db        721 AATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTG 780

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
              | ||| || |  ||    || |   | ||||||  ||| | ||||||| |||| | |||
Db        781 GACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTAT 840

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
                | | |||||  ||||    |  |  |     |     ||  |    ||||| ||  ||
Db        841 CGCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAG 900

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
              |||  |   |  |||||||| || ||  |||| ||||||    |||  |  ||  | ||
Db        901 CGCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTG 960

Qy       4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
              || ||||||| ||||   || |||||||| ||    | | || |  | ||||| ||||||
Db        961 GCCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAAC 1020

Qy       4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
               |    |  ||| | |||||||| ||  | || |||||  ||||||||   | ||||||
Db       1021 GGGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCC 1080

Qy       4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
                 ||||| ||    ||      ||    ||| ||  | ||||| || |||||||| |||
Db       1081 AGCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGC 1140

Qy       4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
                ||| ||  |||| ||| ||   || |  ||| |||    ||| || || || |   |
Db       1141 CGCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAA 1200

Qy       4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
              |||||||  | |  ||| || |   | ||  ||||  |         | | || || |
Db       1201 CTGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGC 1260

Qy       4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
              ||    |||||  ||  |   |||   ||||||||||||   || || ||| | ||||||
Db       1261 CCCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCG 1320

Qy       5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
              ||  ||||||||  |||||||||||||||||||||||||||   ||  |||   ||  ||
Db       1321 ATCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTC 1380

Qy       5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
              ||  | ||||| || || |||||||| || |||||| | || || ||  | ||    |||
Db       1381 TACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCC 1440

Qy       5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCG 5161
              |||||||| ||   | | ||||||  ||| |  ||||
Db       1441 AAGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCG 1477
```

```
RESULT 15
US-09-546-969-6
; Sequence 6, Application US/09546969
; Patent No. 6207409
;  GENERAL INFORMATION:
;    APPLICANT:  Hohmann, Hans-Peter
;    APPLICANT:  Pasamontes, Luis
;    APPLICANT:  Tessier, Michel
;    APPLICANT:  van Loon, Adolphus
;    TITLE OF INVENTION:  FERMENTATIVE CAROTENOID PRODUCTION
;    NUMBER OF SEQUENCES:  47
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hoffmann-La Roche Inc.
;      STREET:  340 Kingsland Street
;      CITY:  Nutley
;      STATE:  NJ
;      COUNTRY:  USA
;      ZIP:  07110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/546,969
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/660,645
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Pokras, Bruce A.
;      REGISTRATION NUMBER:  32,748
;      REFERENCE/DOCKET NUMBER:  RAN 6002/170
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (201) 235-5801
;      TELEFAX:  (201) 235-2363
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1482 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
```

```
US-09-546-969-6

  Query Match             8.8%;  Score 619.4;  DB 3;  Length 1482;
  Best Local Similarity   63.7%;  Pred. No. 1.5e-159;
  Matches  941;  Conservative    0;  Mismatches  536;  Indels    0;  Gaps    0;

Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              ||||    | |    || || |||||||| || ||||| || ||||| || || ||||||
Db          1 ATGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTG 60

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              ||| |||| |||||  | ||||||| |  ||||||   ||||||||| || ||| ||||||
Db         61 CAATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCC 120

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              ||||| |   | ||    |||     |||||||| || || |||||  | ||||| |||
Db        121 TATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCC 180

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
                |  | |   |||| |||   | ||    ||  || || |      | ||||   ||||
Db        181 GACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTG 240

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
               |||||||  |||||||||| || |||   ||||  || |||       | || ||||
Db        241 CTGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTG 300

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              ||  |  |      ||||   |  ||| ||||| | ||||||||    |||||| || |||
Db        301 AACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGC 360

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              ||||| |  || |  |  ||  |    ||   || | | | |||| ||||||||| || ||
Db        361 TATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGG 420

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              ||   ||| || ||| || | |  | |||||  || |||||  | ||   || |||
Db        421 ACCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTG 480

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              || ||||  |||   ||| ||||||||||||||||   || || ||||||   |||||||| 
Db        481 CAGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGG 540

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||| || || |||||| |  ||||||| ||||| || || ||| | ||    ||||||
Db        541 CAGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATC 600

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
              ||| |  | || |||||||||||| ||     | ||||| ||||  |    |||||||||
Db        601 TATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACC 660

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
                |  ||||||   ||||||||     |||||||||    | ||||||  |||| | |||
Db        661 AACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTG 720

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
              ||||||   || |  | | | || ||    ||| |||||   |  |||||   | |   |
Db        721 AATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTG 780

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
              | ||| || |  ||    ||  | |||  ||| ||||||| |||||||| |||| ||| 
Db        781 GACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTAT 840

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
                | | |||||  ||||    |  |  |     |     ||  |    ||||| ||  ||
Db        841 CGCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAG 900

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
              |||  |   |  |||||||| || ||  |||| ||||||    |||  |  ||  | || 
Db        901 CGCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTG 960

Qy       4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
              || ||||||| ||||   || |||||||| ||    | | || |  | ||||| ||||||
Db        961 GCCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAAC 1020

Qy       4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
               |    |  ||| | |||||||| ||  |  || |||||  ||||||||   | |||||| 
Db       1021 GGGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCC 1080

Qy       4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
                 ||||| ||    ||      ||    ||| ||  | ||||| || |||||||| |||
Db       1081 AGCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGC 1140

Qy       4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
                ||| ||  |||| ||| ||   || |  ||| |||    ||| || || || |   | 
Db       1141 CGCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAA 1200

Qy       4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
              |||||||  | |  ||| || |   | ||  ||||  |         | | || || |  
Db       1201 CTGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGC 1260

Qy       4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
              ||    |||||  ||  |  | |||   ||||||||||||   || || || | ||||||
Db       1261 CCCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCG 1320

Qy       5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
              ||  ||||||||  ||||||||||||||||||||||||||||||   ||  |||   ||  ||
Db       1321 ATCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTC 1380

Qy       5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
              ||  | ||||| || || |||||| || |||||| | || || ||  |  ||    |||
Db       1381 TACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCC 1440

Qy       5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCG 5161
              |||||||| ||   | | ||||||  ||| |  ||||
Db       1441 AAGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCG 1477

Search completed: January 25, 2005, 18:28:48
Job time : 509 secs
```

```
                     GenCore version 5.1.6
              Copyright (c) 1993 - 2005  Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:         January 25, 2005, 10:29:07 ; Search time 3350 Seconds
                                              (without alignments)
                                              12004.619 Million cell updates/sec

Title:          US-10-808-979-18
Perfect score:  6999
Sequence:       1 gcgatggcaatggaaaatgt..........gacgttttagacatgcaata 6999

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters:      8600550

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
                10:  /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
                12:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                13:  /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
                14:  /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
                15:  /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
                16:  /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq:*
                17:  /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
                18:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                19:  /cgn2_6/ptodata/1/pubpna/US11_NEW_PUB.seq:*
                20:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                21:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.

                                  SUMMARIES
                 %
Result          Query
  No.   Score   Match Length DB  ID                         Description
-------------------------------------------------------------------------------
    1    6999   100.0   6999  18  US-10-808-979-18            Sequence 18, Appl
    2  4326.2    61.8   8675  18  US-10-804-677-18            Sequence 18, Appl
    3  2619.6    37.4   8609  18  US-10-735-019-28            Sequence 28, Appl
    4  2619.6    37.4   8609  18  US-10-734-778-40            Sequence 40, Appl
    5  2619.6    37.4   8609  18  US-10-734-936-37            Sequence 37, Appl
    6  2619.6    37.4   8609  18  US-10-735-442-64            Sequence 64, Appl
    7  2138.6    30.6   8814  18  US-10-808-807-18            Sequence 18, Appl
    8  2135.2    30.5  12753  16  US-10-041-018-19            Sequence 19, Appl
    9  1618.8    23.1   7494  17  US-10-474-536-45            Sequence 45, Appl
   10  1618.8    23.1   8547  17  US-10-474-536-46            Sequence 46, Appl
   11    1482    21.2   1482  18  US-10-808-979-7             Sequence 7, Appli
   12    1278    18.3   1278  18  US-10-808-979-3             Sequence 3, Appli
   13    1167    16.7   1167  18  US-10-808-979-5             Sequence 5, Appli
   14  1085.2    15.5   1482  18  US-10-804-677-7             Sequence 7, Appli
c  15  1035.4    14.8   8625  10  US-09-920-923-1             Sequence 1, Appli
c  16  1035.4    14.8   8625  16  US-10-695-980-1             Sequence 1, Appli
c  17  1035.4    14.8  11233  10  US-09-920-923-27            Sequence 27, Appl
c  18  1035.4    14.8  11233  16  US-10-695-980-27            Sequence 27, Appl
   19     930    13.3    930  18  US-10-808-979-9             Sequence 9, Appli
   20     906    12.9    906  18  US-10-808-979-1             Sequence 1, Appli
   21   875.4    12.5   1278  18  US-10-804-677-3             Sequence 3, Appli
   22   861.4    12.3   1479  18  US-10-808-807-7             Sequence 7, Appli
   23   808.6    11.6   1479  10  US-09-941-947A-31           Sequence 31, Appl
   24   808.6    11.6   1479  15  US-10-218-118-7             Sequence 7, Appli
   25   808.6    11.6   1479  16  US-10-363-567-31            Sequence 31, Appl
   26   808.6    11.6   1479  16  US-10-466-656-7             Sequence 7, Appli
   27   808.6    11.6   1479  18  US-10-735-019-7             Sequence 7, Appli
   28   808.6    11.6   1479  18  US-10-734-778-7             Sequence 7, Appli
   29   808.6    11.6   1479  18  US-10-735-442-7             Sequence 7, Appli
   30   808.6    11.6   1479  18  US-10-848-307-7             Sequence 7, Appli
   31     695     9.9   1167  18  US-10-804-677-5             Sequence 5, Appli
   32   692.6     9.9   1521  14  US-10-166-037-5             Sequence 5, Appli
   33   630.8     9.0    930  18  US-10-804-677-9             Sequence 9, Appli
   34   629.2     9.0    906  18  US-10-804-677-1             Sequence 1, Appli
   35   619.4     8.8   1482  9  US-09-547-267-6             Sequence 6, Appli
   36   535.6     7.7   1473  15  US-10-369-493-23921         Sequence 23921, A
c  37     534     7.6    534  18  US-10-808-979-11            Sequence 11, Appl
   38   510.6     7.3   1518  15  US-10-369-493-41517         Sequence 41517, A
   39   480.8     6.9   1494  15  US-10-369-493-34266         Sequence 34266, A
   40   449.2     6.4   1296  10  US-09-941-947A-27           Sequence 27, Appl
   41   449.2     6.4   1296  15  US-10-218-118-3             Sequence 3, Appli
   42   449.2     6.4   1296  16  US-10-363-567-27            Sequence 27, Appl
   43   449.2     6.4   1296  16  US-10-466-656-1             Sequence 1, Appli
   44   449.2     6.4   1296  18  US-10-735-019-3             Sequence 3, Appli
   45   449.2     6.4   1296  18  US-10-734-778-3             Sequence 3, Appli

                                  ALIGNMENTS

RESULT 1
US-10-808-979-18
; Sequence 18, Application US/10808979
; Publication No. US20040268439A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2360 US NA
;  CURRENT APPLICATION NUMBER: US/10/808,979
;  CURRENT FILING DATE:  2004-03-25
;  PRIOR APPLICATION NUMBER: US 60/471,904
;  PRIOR FILING DATE: 2003-05-20
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 6999
;   TYPE: DNA
;   ORGANISM: Enterobacteriaceae strain DC260
US-10-808-979-18

  Query Match             100.0%;  Score 6999;  DB 18;  Length 6999;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6999;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 GCGATGGCAATGGAAAATGTTGCGTGAGCTTTTCGTCTAACTGCGGCATCAGCGGCTGAA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GCGATGGCAATGGAAAATGTTGCGTGAGCTTTTCGTCTAACTGCGGCATCAGCGGCTGAA 60

Qy         61 TAATCAGCACGGCAGGTTTGGTATTCGTCATTTTATTGTCCATTAGCGGTATTTATTTAC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TAATCAGCACGGCAGGTTTGGTATTCGTCATTTTATTGTCCATTAGCGGTATTTATTTAC 120
```

```
Qy        121 CACCAGCCTAACGGAGTTATTTATGTTACGGCGTTGCTGTTACTTATTTCGCTAATAAGA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CACCAGCCTAACGGAGTTATTTATGTTACGGCGTTGCTGTTACTTATTTCGCTAATAAGA 180

Qy        181 TCACGCATAGCATTATTAACAATATTTACCTGGTGCGCATGAATACGCACCCTACAAAGT 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TCACGCATAGCATTATTAACAATATTTACCTGGTGCGCATGAATACGCACCCTACAAAGT 240

Qy        241 CAAGTCCCTCGCTGGCGAACTCACCTTACGCAGTCTACGGTTAATCAAAAAGCATAAAAA 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CAAGTCCCTCGCTGGCGAACTCACCTTACGCAGTCTACGGTTAATCAAAAAGCATAAAAA 300

Qy        301 TTTCACCAACCATGGATAGCCATTATGACCACCCATGTCGACACCACAGCACATCAGACA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TTTCACCAACCATGGATAGCCATTATGACCACCCATGTCGACACCACAGCACATCAGACA 360

Qy        361 AGCGAACTCCTTCAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 AGCGAACTCCTTCAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCC 420

Qy        421 GGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAA 480

Qy        481 CGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAA 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAA 540

Qy        541 AATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTG 600

Qy        601 GATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGC 660

Qy        661 GAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAA 720

Qy        721 GTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTC 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTC 780

Qy        781 TCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGC 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGC 840

Qy        841 ACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTG 900

Qy        901 TTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAA 960

Qy        961 AGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCC 1020

Qy       1021 GACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTA 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTA 1080

Qy       1081 CAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGAT 1140

Qy       1141 GCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACTCGCCAATATATGCACGCGCTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACTCGCCAATATATGCACGCGCTG 1200
```

```
Qy       1201 TTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 TTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGG 1260

Qy       1261 TGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 TGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGC 1320

Qy       1321 TGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCG 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 TGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCG 1380

Qy       1381 ACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 ACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGC 1440

Qy       1441 CCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATC 1500

Qy       1501 TCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACA 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 TCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACA 1560

Qy       1561 TCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGC 1620

Qy       1621 ATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGC 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 ATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGC 1680

Qy       1681 TTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 TTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCA 1740

Qy       1741 GCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGC 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGC 1800

Qy       1801 GGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAA 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 GGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAA 1860

Qy       1861 TCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACG 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACG 1920

Qy       1921 CCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGC 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 CCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGC 1980

Qy       1981 TGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGT 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 TGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGT 2040

Qy       2041 TTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTG 2100

Qy       2101 GGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGG 2160

Qy       2161 ATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 ATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGT 2220

Qy       2221 TAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTT 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 TAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTT 2280

Qy       2281 TTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCAC 2340
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 TTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCAC 2340

Qy       2341 GCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTT 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 GCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTT 2400

Qy       2401 ACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTG 2460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2401 ACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTG 2460

Qy       2461 CCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCT 2520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2461 CCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCT 2520

Qy       2521 GATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGC 2580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2521 GATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGC 2580

Qy       2581 GCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCA 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 GCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCA 2640

Qy       2641 TCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCG 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 TCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCG 2700

Qy       2701 CTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCT 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 CTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCT 2760

Qy       2761 GCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCT 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 GCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCT 2820

Qy       2821 TTACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCAC 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 TTACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCAC 2880

Qy       2881 GCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGG 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 GCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGG 2940

Qy       2941 CCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCA 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCA 3000

Qy       3001 GCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCA 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 GCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCA 3060

Qy       3061 AGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 AGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGA 3120

Qy       3121 TACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGA 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 TACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGA 3180

Qy       3181 CTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATT 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 CTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATT 3240

Qy       3241 ACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTG 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 ACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTG 3300

Qy       3301 CAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGT 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGT 3360

Qy       3361 GGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACA 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 GGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACA 3420

Qy       3421 TATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCG 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 TATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCG 3480

Qy       3481 CATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 CATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCG 3540

Qy       3541 GCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAAC 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 GCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAAC 3600

Qy       3601 GCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAA 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 GCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAA 3660

Qy       3661 TTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGC 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 TTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGC 3720

Qy       3721 GGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 GGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAG 3780

Qy       3781 CGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCC 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 CGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCC 3840

Qy       3841 GGACCCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAA 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 GGACCCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAA 3900

Qy       3901 TCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAA 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 TCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAA 3960

Qy       3961 GATGGCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACG 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 GATGGCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACG 4020

Qy       4021 TTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTT 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 TTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTT 4080

Qy       4081 AGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGC 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 AGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGC 4140

Qy       4141 GTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAA 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 GTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAA 4200

Qy       4201 TTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGT 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 TTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGT 4260

Qy       4261 AATCCTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGC 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 AATCCTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGC 4320

Qy       4321 GTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAG 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAG 4380

Qy       4381 GACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAAT 4440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4381 GACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAAT 4440

Qy       4441 CGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCC 4500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4441 CGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCC 4500
```

```
Qy       4501 AATGCCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAA 4560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4501 AATGCCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAA 4560

Qy       4561 CGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGC 4620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4561 CGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGC 4620

Qy       4621 CTGAATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGT 4680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4621 CTGAATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGT 4680

Qy       4681 GAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTG 4740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4681 GAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTG 4740

Qy       4741 CACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTG 4800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4741 CACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTG 4800

Qy       4801 TTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGC 4860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4801 TTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGC 4860

Qy       4861 TTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAA 4920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4861 TTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAA 4920

Qy       4921 TTAGTGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCAC 4980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4921 TTAGTGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCAC 4980

Qy       4981 GGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAAC 5040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4981 GGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAAC 5040

Qy       5041 CGCGATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGC 5100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5041 CGCGATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGC 5100

Qy       5101 GTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCC 5160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5101 GTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCC 5160

Qy       5161 GAATGAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGA 5220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5161 GAATGAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGA 5220

Qy       5221 GTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGT 5280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5221 GTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGT 5280

Qy       5281 ATGCGTGGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCA 5340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5281 ATGCGTGGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCA 5340

Qy       5341 CGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCC 5400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5341 CGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCC 5400

Qy       5401 GCGCCTACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGA 5460
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5401 GCGCCTACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGA 5460

Qy       5461 TCATTCACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATG 5520
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5461 TCATTCACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATG 5520

Qy       5521 CACGCAACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGG 5580
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5521 CACGCAACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGG 5580
```

```
Qy       5581 GCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATC 5640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5581 GCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATC 5640

Qy       5641 ACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAG 5700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5641 ACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAG 5700

Qy       5701 ATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCG 5760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5701 ATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCG 5760

Qy       5761 CCGATACGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAG 5820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5761 CCGATACGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAG 5820

Qy       5821 TGGCGGAGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGC 5880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5821 TGGCGGAGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGC 5880

Qy       5881 GCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTC 5940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5881 GCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTC 5940

Qy       5941 AGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGG 6000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5941 AGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGG 6000

Qy       6001 CGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGC 6060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6001 CGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGC 6060

Qy       6061 GTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTACGTCCGTGACGCTGGCGCAGCGT 6120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6061 GTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTACGTCCGTGACGCTGGCGCAGCGT 6120

Qy       6121 GGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACG 6180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6121 GGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACG 6180

Qy       6181 CCCGCGCACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCG 6240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6181 CCCGCGCACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCG 6240

Qy       6241 CGGGATATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAG 6300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6241 CGGGATATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAG 6300

Qy       6301 CGCGCCGTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACC 6360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6301 CGCGCCGTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACC 6360

Qy       6361 GACATAAATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTC 6420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6361 GACATAAATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTC 6420

Qy       6421 AAACTTACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCAT 6480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6421 AAACTTACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCAT 6480

Qy       6481 GATGTATTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGAT 6540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6481 GATGTATTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGAT 6540

Qy       6541 AAGCACGTTCCATAACCAGAGCATTGTTCGTCCATTTGTGGAAAAGGGAAGTACTAAAGG 6600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6541 AAGCACGTTCCATAACCAGAGCATTGTTCGTCCATTTGTGGAAAAGGGAAGTACTAAAGG 6600

Qy       6601 TGGACGCGGATGAGTGATGGCGCAAGGTTTACCATGTTTAGAAATTTTAAAAGTCCATAA 6660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6601 TGGACGCGGATGAGTGATGGCGCAAGGTTTACCATGTTTAGAAATTTTAAAAGTCCATAA 6660

Qy       6661 CACGTTATGAACGCTGCATTGCAGAAAGCGCAGATTTCACACATACTCACCACACTTATC 6720
```

```
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6661 CACGTTATGAACGCTGCATTGCAGAAAGCGCAGATTTCACACATACTCACCACACTTATC 6720

Qy        6721 AATACACGTGTTAACTACATGGGGATTTATGCCTTCTACAGCCGTAAGACAAAAAAAAA 6780
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6721 AATACACGTGTTAACTACATGGGGATTTATGCCTTCTACAGCCGTAAGACAAAAAAAAA 6780

Qy        6781 CTGTCAGTGTGACACTTGAACCTGCTCTACTCGAGCAAGCCAGAGAGGCAGGGCTCAATT 6840
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6781 CTGTCAGTGTGACACTTGAACCTGCTCTACTCGAGCAAGCCAGAGAGGCAGGGCTCAATT 6840

Qy        6841 TATCCGCCATCCTATCCAAAGCTTTGCAACATGAAATTCGCACGACTGCAGCAGAAAGAT 6900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6841 TATCCGCCATCCTATCCAAAGCTTTGCAACATGAAATTCGCACGACTGCAGCAGAAAGAT 6900

Qy        6901 GGAAGCGTGAAAACAGTGAAGGTTTGCAGGAACTCAATCGCATAACCGAAGAGCACGGTT 6960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        6901 GGAAGCGTGAAAACAGTGAAGGTTTGCAGGAACTCAATCGCATAACCGAAGAGCACGGTT 6960

Qy        6961 TATTGTCGGATGAATACAGGACGTTTTAGACATGCAATA 6999
               |||||||||||||||||||||||||||||||||||||||
Db        6961 TATTGTCGGATGAATACAGGACGTTTTAGACATGCAATA 6999


RESULT 2
US-10-804-677.-18
; Sequence 18, Application US/10804677
; Publication No. US20040224383A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  APPLICANT:  Sedkova, Natalia
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2346 US NA
;  CURRENT APPLICATION NUMBER: US/10/804,677
;  CURRENT FILING DATE:  2004-03-19
;  PRIOR APPLICATION NUMBER: US 60/468,596
;  PRIOR FILING DATE: 2003-05-07
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 8675
;   TYPE: DNA
;   ORGANISM: Pectobacterium cypripedii DC416
US-10-804-677-18

  Query Match             61.8%;  Score 4326.2;  DB 18;  Length 8675;
  Best Local Similarity   80.1%;  Pred. No. 0;
  Matches 5122;  Conservative    0;  Mismatches 1268;  Indels    5;  Gaps    3;

Qy         236 AAAGTCAAGTCCCTCGCTGGCGAACTCACCTTACGCAGTCTACGGTTAATC-AAAAAGCA 294
               ||||||   || |||||||||| |||| ||||||   |||||||||||||  ||||||||
Db        1119 AAAGTCGCTTCGCTCGCTGGCGCACTCCCCTTACCGGGTCTACGGTTAATTGAAAAAGCA 1178

Qy         295 TAAAAATTTCACCAACCATGGATAGCCATTATGACCACCCATGTCGACACCACAGCACAT 354
                || ||||| || ||||||||| ||||  ||||||| |||||||||| |||||||||
Db        1179 CAAGAATTTAACTAACCATGGAAAGCCGCTATGACCGCCCATGTCGATACCACAGCAAGC 1238

Qy         355 CAGACAAGCGAACTCCTTCAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTG 414
               |||  |||||| ||||||||| |||| ||    || ||||| ||||||||||||||| ||
Db        1239 CAGGAAAGCGATCTCCTTCAGTTGCATCACGCATTGCAGGCCCATCTTGAACATTTATTG 1298

Qy         415 CCTGCCGGACAGCAAAGCGATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAG 474
               |||||||| |||||   ||||||||| || ||||| ||||||||||| |||||||| | |
Db        1299 CCTGCCGGGCAGCAGGCCGATCGCGTTCGGGCCGCCATGCGTGCCGGCACGCTGGCACCG 1358

Qy         475 GGCAAACGTATTCGTCCTTTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTG 534
               |||||||||||||||||  | || |||||||||||||| ||||||||||| || ||  ||
Db        1359 GGCAAACGTATTCGTCCGCTCTTGCTGCTGCTGGCAGCACGCGATATGGGCTGTGACGTG 1418

Qy         535 ACGCAAAATGGCGTTCTCGATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTG 594
                ||||  | ||| | || ||||| |||||||| || |||||||||||||| || || |||
Db        1419 GCGCAGCAGGGCATCCTTGATCTTGCCTGTGCGGTCGAAATGGTGCACGCTGCCTCACTG 1478

Qy         595 ATTCTGGATGACATTCCCTCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTG 654
               || || || |||||||| || ||||||||||| | |||||| ||||| |||||  |  |
Db        1479 ATCCTCGACGACATTCCATCAATGGATAACGCCCGGATGCGACGTGGGCGCCCGGCAATC 1538

Qy         655 CATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCA 714
               ||  | |||| ||| |||||||||||||| || || ||  ||||||| || ||||||||
Db        1539 CACTGTGAATATGGGGAAAACGTGGCGATCCTGGCAGCGGTCGCGCTACTCAGCCGCGCC 1598

Qy         715 TTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCT 774
               ||||| |||||||||| | || || ||| |||| || |  || ||| | ||||| |||||
Db        1599 TTTGAGGTGATTGCCCTCGCGCCGGGTCTGCCAGCAACGCACAAAGCCGAAGCCATTGCC 1658

Qy         775 GAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCAC 834
               || |||||| ||||||| |||||||||||  | || || || || ||||||||||||||
Db        1659 GAGCTCTCCTCTGCCGTGGGCCTGCAGGGACTGGTTCAGGGTCAGTTCCAGGATCTGCAT 1718

Qy         835 GACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGC 894
               |||||| | || |||||||| |||||||| ||| || ||||||| |||||||||||||||
Db        1719 GACGGCGCACACAGCCGCAGTCCGGAAGCCATCACCCTGACCAATGAACTGAAAACCAGC 1778

Qy         895 GTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTG 954
               || |||||||||||||||||||| |||||||||||||| || || |||  |||||||
Db        1779 GTCCTGTTTCGCGCCACGCTGCAGATGGCGGCGATTGCGGCCGATGCGTCAGTGCAGGTA 1838

Qy         955 CGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGAC 1014
               || ||  |  | ||||  || || |||||||| || ||||| || ||  | || ||||||
Db        1839 CGTCAGCGTTTAAGCTATTTTGCGCAGGATTTAGGTCAGGCTTTCCAGTTACTGGACGAC 1898

Qy        1015 CTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACG 1074
               || || || || |  || |||||||| || ||    || |||||||||||||||||||||
Db        1899 CTGGCGGATGGCTCTAAGCACACCGGCAAGGACTGTCATCAGGATCAGGGCAAATCCACG 1958

Qy        1075 CTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGC 1134
               ||||| ||||||||| ||  | || || || || ||||| |||||||| || ||  |||||
Db        1959 CTGGTGCAGATGCTGGGCCCGGAAGGGGCTGAGCGTCGTCTGCGCGACCATCTAAGCAGC 2018

Qy        1135 GCAGATGCACACCTTGCCTGCGCCTGCCATCGCGGCATCGCCACTCGCCAATATATGCAC 1194
               || ||||||||||||||||||||||||||||||||  ||||||| || ||||||||||||
Db        2019 GCCGATGCACACCTTGCCTGCGCCTGCCATCGCGGTGTCGCCACCCGTCAATATATGCAC 2078

Qy        1195 GCGCTGTTTAATCAACAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACT 1254
               || ||||||||||||||||| ||||| |||||||||   |||     ||  |||| || |
Db        2079 GCCCTGTTTAATCAACAGCTGGCGATGTTCAACTGAAGCCGGTCATACCTATGGGGCATT 2138

Qy        1255 TTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAA 1314
               |||| || || ||||| ||||||||||||||||||||||| ||||||||| | || ||||
Db        2139 TTGCCGTTATTGCGCCACCGCTCTACAGCCACTTTCACGCATTGCAGGCGCTGGCGCAAA 2198

Qy        1315 CGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGC 1374
               |||||||||||||||| ||||||||||| || |||||||||   ||||| ||||| ||||
Db        2199 CGCTGCTGGCGCGCGGACATCGCATCACCTTTATCCAGCAAAGTGATGCACGCACCTTGC 2258

Qy        1375 TTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGC 1434
               | |||||||| ||||| |  ||||| ||||||||| | |  |||||||||||||| ||||
Db        2259 TGAGCGACGAGCGCATTGCCTTTGTGGCCGTCGGCGAGCGCACGCATCCTGCCGGATCGC 2318

Qy        1435 TGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCG 1494
               |  |   ||   |    ||||||||| | |||||||| ||||||||||||||||||||
Db        2319 TCTCCAGCGAACTCAGGCGGCTGGCCGCACCGGGCGGGCTGTCGCTGTTTCGCGTGATTC 2378

Qy        1495 ACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCAT 1554
               ||||||| ||   |  |||||||||||| |||||||||||||| ||||| ||| ||
Db        2379 ACGATCTGGCCAGCACCACCGATATGCTATGCCGCGAACTGCCCGCGGTGCTGCAACGGC 2438

Qy        1555 TGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAG 1614
               || |  ||||||||||||| |||||  ||||||||||||| || ||  |||| || || |
Db        2439 TGCAGGTCGATGGCGTGATTGCCGATCAAATGGAAGCGGCTGGTGGTCTGGTGGCAGAGG 2498

Qy        1615 CGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGA 1674
```

```
              || | || |||||||| || ||||||||||||||  ||||||||||||| ||||| |  |
Db       2499 CGTTACAGCTGCCGTTCGTGTCGGTGGCCTGCGCGCTGCCGGTCAATCGCGAAGCGGCCA 2558

Qy       1675 TTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTC 1734
              ||||||| | |||||||||||| || ||||| |||||||| ||||||||| | || | ||
Db       2559 TTCCGCTGGTGGTGATGCCCTTTCGCTTTGCTCAGGATGAGAAAGCGCTGCAGCGCTATC 2618

Qy       1735 AGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAAC 1794
              |||||||||| || ||||| || |||||||||||||||||| ||||  || |||      |
Db       2619 AGGCCAGCAGTGACATCTACGACCGCATCATGCGTCGTCATGGCGCTGTCATCGCTCGTC 2678

Qy       1795 ACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGG 1854
              | ||||| || ||     || | || || |  || ||||||||||| |||||||||||||
Db       2679 ATGCGCGCGCCTTCGGCCTGCCCGAACGCCATGGCTTACATCAGTGTCTGTCGCCGCTGG 2738

Qy       1855 CACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCT 1914
              | |||||||| ||| |||||||| || ||||||||||||| ||||||||||| |||||||
Db       2739 CGCAAATCAGTCAGCTGGTGCCCGCTTTTGATTTTCCACGCCAGCAACTGCCAGCCTGCT 2798

Qy       1915 ATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGC 1974
              |||||  |||||| || || ||  | || ||| ||   |  |||||   || |||||||
Db       2799 ATCACAGCGTGGGTCCGCTGCGGACTCCAGTTGCTAGCGGCGCGCTCGCCGCACCCTGGC 2858

Qy       1975 CAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCC 2034
              ||||||||||| ||||||||||| ||||||||||||||| |||||| || || |||||||| |
Db       2859 CAGCGCTGCGCCAGCCGGTGGTGTATGCCTCGCTGGGCACGCTACAGGGGCATCGCTTTC 2918

Qy       2035 GGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCC 2094
              | ||||||||||||||||| ||||| |||||| | | || ||| |||||||||| | || |
Db       2919 GCCTGTTTCTGCATCTGGCTCAGGCCTGCCGCAATCAGCAGCTGTCGCTGGTGGTGGCAC 2978

Qy       2095 ATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGG 2154
              | ||||| || || | ||||   ||| | ||||||||||    || || || || |||||||
Db       2979 ACTGTGGCGGGTTGACCGCCAGCCAGGCACATCAGCTCAGACTGGCCGGTGCTGCGTGGG 3038

Qy       2155 TGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATG 2214
              |||| ||||| || |||||||| || || || ||||| ||||| |||||||||||||||| |
Db       3039 TGACCGATTTTGTGGATCAGCGGGCGGCGCTGCAGCATGCGCAACTGTTTATCACTCACG 3098

Qy       2215 CCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGA 2274
              ||||  | ||||| |||||||||||||||||| || || ||||||||||||||||||||||
Db       3099 CCGGTCTGAACAGTGCGCTGGAAGCACTGGAGTGTGGCACGCCGATGCTGGCGCTGCCGA 3158

Qy       2275 TTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCG 2334
              | || || |||||||||||||||||||||| | ||||||||||| | ||| || || ||||
Db       3159 TCGCCTTCGATCAGCCCGGCGTGGCGGCACGTATTGAGTGGCACGGCGTCGGCCGGCGCG 3218

Qy       2335 CATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACG 2394
              | ||||| || ||||| ||    || |||||||| || ||||||||| ||||||  ||||
Db       3219 CCTCACGTTTCAGCCGGGTCGCGCAGCTGGAGCACCACCTGCAACAGTTGCTGAGTGACG 3278

Qy       2395 ATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCC 2454
              |||| ||    || || ||||||||| ||||||||||||||||||||||| || || || ||
Db       3279 ATCGCTATCGTCTGCGCATGTCAGCCATTCAGGCGCAGCTGCAGCGGGCCGGTGGCTGTA 3338

Qy       2455 AGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGG 2514
               ||| || || || || |||||||||||||||||| |||||||||| ||||||||||||||||
Db       3339 CGCGCGCGGCTGATATTGTCGAGCAGGCGCTGTGTCAGCAGCAAATCGTGCTGGCGGAGG 3398

Qy       2515 CGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTT 2574
              | |||||||||||| | |  || ||||| || | |||||||||| || |||||| || ||  |
Db       3399 CCACCTGATGCGCGCACCTTATGATGTCATTCTGGTCGGTGCCGGCCTGGCTAACGGGCT 3458

Qy       2575 GATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGA 2634
              ||||||||||||| | || ||  |||||||   ||| ||     || | |||||||  |
Db       3459 GATTGCGCTGCGTTTACGCCAGCTGCAGCCCGCACTTAAGGTTTTGCTACTGGAGAGTCA 3518

Qy       2635 TGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACA 2694
               ||||| ||||| |||||||||||||||||||||| |||| |   ||  ||||||  |  ||
Db       3519 GGCGCAGCCGGCCGGCAATCATACCTGGTCGTTCCATCGCGAAGACGTCAGCGAAGCGCA 3578

Qy       2695 ACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCC 2754
                ||||||||||  | |||||| || | | ||| ||| | ||||||||||| || || ||
Db       3579 GTTTCGCTGGCTCGAGCCGCTGCTTTCGGCGCGCTGGCCCGGTTATCAGGTACGCTTCCC 3638

Qy       2755 TGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCG 2814
                | ||||| ||| | |||| || || ||| || || || || |  ||||||||||||
Db       3639 CACCCTGCGTCGCCAGCTGGATGGTGAATATTGCTCGATTGCCTCGGAGGATTTTGCCCG 3698

Qy       2815 CCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAA 2874
               ||  |  |   || | | ||| ||  ||  | ||     |  | ||     ||||
Db       3699 GCACTTACAGCAGGTGCTCGGTGCCGCGCTACGCACCGCAGCGCCGGTCAGCGAGGTCTC 3758

Qy       2875 ACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGG 2934
              ||||||   ||| |  |||||||||||| |     |     ||||| | ||||||| ||
Db       3759 ACCCACCGGGGTCAGACTGGCGGATGGCGGGATGTTACAGGCGCAGGCGGTGATTGACGG 3818

Qy       2935 TCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGA 2994
               ||||| |||||||||||| || |||||||||||| || ||||||||  ||| | || || ||
Db       3819 ACGCGGGCTGCAGCCGACACCGCATCTGCAGCTCGGCTATCAGGCATTTGTCGGTCAGGA 3878

Qy       2995 GTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGA 3054
              |||||| |||||   |||||| |||||||||||||||| ||  ||||||| |||| ||||||
Db       3879 GTGGCAACTGGCCGCGCCGCATGGCCTGCAGCAGCCAATATTGATGGACGCCAGCGTCGA 3938

Qy       3055 TCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGAT 3114
              ||||||   |||||||||| ||||| ||||| ||||||||||| |||   ||  || ||||
Db       3939 TCAGCAGCAGGGTTATCGCTTTGTTTACACCCTGCCGCTCAGTGCCAGCCGTTTACTGAT 3998

Qy       3115 TGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACAT 3174
              |||||||||||| ||| | ||||  || ||||||  |        || || |  |||||
Db       3999 TGAAGATACCCACTACATCAACCATGCCACGCTGGATGCCGCACAGGCGCGCCGTCACAT 4058

Qy       3175 CGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGG 3234
                | || |||||| | ||||  |||||||   || |   |||||||||| || ||||||||
Db       4059 TACGGATTATGCCCACCAGCGCGGCTGGAATTTGCGCCAGCTGCTGCGCGAGGAGCACGG 4118

Qy       3235 CATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCA 3294
              |    | ||||| || || |||||| | |||||||  |||||||||||||||||| ||| ||
Db       4119 CTCGCTGCCGATCACGCTCAGCGGCGATATCGATCAGTTCTGGCAACAGCAGCACGGGCA 4178

Qy       3295 AGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTC 3354
              | |||||||||| ||||||||||||| |||||||||| ||||||||||||||||||  ||||  |
Db       4179 ACCGTGCAGCGGGCTGCGCGCCGGACTGTTTCACGCCACCACCGGTTACTCGCTGCCCGC 4238

Qy       3355 CGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAG 3414
              ||| |||||||| ||||||  | | ||      | ||||| |||| || | |||||| ||
Db       4239 CGCGGTGGCGCTGGCGGAGAAGATTGCCAGCACGCTGCCCGCCGACGCTCACACGCTGAG 4298

Qy       3415 CCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCT 3474
              |||    ||||||  ||||||||||||||| |||||||| ||||| |||||||||||||| |
Db       4299 CCACTGCATCGAATCCTTTGCCCGTCAGCACTGGCGCGAGCAGCGCTTTTTCCGTCTGTT 4358

Qy       3475 AAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTT 3534
              ||| |||||||||||| | |||||   |||  | |||||||||||||||| | |||| |||||
Db       4359 AAATCGCATGCTGTTTCTTGCCGGACGGCCTGAACAGCGCTGGCGCGTAATGCAGCGTTT 4418

Qy       3535 TTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGA 3594
              ||||||||| || ||||| || ||||||||||||||||||||||||||||||||||  ||||
Db       4419 TTACCGGCTTGACGCCGGATTGATTAGCCGCTTTTACGCCGGGCAACTGCGCCTCAGCGA 4478

Qy       3595 TAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGC--- 3651
              |||| | || ||||||||||||||| ||||||||| ||  |||| ||||||||||||||
Db       4479 TAAAGCACGCATTCTGTGCGGCAAACCGCCGGTCCCTCTCGGCGAAGCGCTGCGCGCATT 4538

Qy       3652 GCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCA 3711
              | || |||  |||     ||||||||||||| |||||||||||||| ||||||||||||||||
Db       4539 GATGATGACCTCTCCGTTACCAGGGAAGAAATAATGAAACGCACCTATGTGATTGGCGCA 4598

Qy       3712 GGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTA 3771
              ||||| || |||||||||||||||||||||| ||||||||||| ||| | ||   |||  |
Db       4599 GGCTTCGGTGGCCTGGCGCTGGCGATTCGTCTGCAAGCGGCCGGCGTGCCGGTCACGCTG 4658

Qy       3772 CTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACC 3831
              || || |||||||| || || ||||| |||||||||||||| | ||||     || ||||||
Db       4659 CTGGAACAGCGCGATAAGCCTGGCGGGCGCGCCTATGTGTATCAGGATCAGGGTTTTACC 4718
```

```
Qy       3832 TTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTG 3891
              || |||||||| || |||||||| |||||||||||||| ||||| | | |||| ||||||
Db       4719 TTTGATGCCGGTCCGACGGTGATTACCGATCCCAGCGCTATCGAGGCGCTGTTTACGCTG 4778

Qy       3892 GCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTG 3951
              ||||| ||    ||||| ||||| ||||| ||||||||||| ||||| || |||||||||
Db       4779 GCAGGCAAGCAACTCAGTGATTATGTCGACCTGATGCCGGTGACGCCATTTTATCGCCTG 4838

Qy       3952 TGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAG 4011
              ||||||||||| ||||  ||||| || |||||||| |||||| ||| |||||||||||||
Db       4839 TGCTGGGAAGACGGCAGGCAGCTGGACTACGACAACAATCAGGCGCAGCTGGAGCAGCAG 4898

Qy       4012 ATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGT 4071
              || ||||| || ||||| || ||||| |  || || || || ||||| ||||||||||
Db       4899 ATTGCCACTTTTAATCCCCAGGATGTCGCCGGTTACCGCCAGTTTCTGGCCTATTCACAG 4958

Qy       4072 GAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGAC 4131
              || || ||| | ||||| ||||||||||| ||||| || || ||||||||  | || |||
Db       4959 GATGTGTTTCGTGAGGGCTATCTGAAACTGGGCACCGTACCTTTTCTGCATTTCCGCGAC 5018

Qy       4132 ATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATG 4191
              |||||||| | || ||| ||| |||| || ||||| || |||||||| ||||||||||||
Db       5019 ATGCTGCGTGCCGGGCCACAGCTGGGTCGGCTGCAGGCCTGGCGCAGTGTCTACAGCATG 5078

Qy       4192 GTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTG 4251
              |||||||||||||||||| |||||||||||||| ||||| |||||||| ||||| ||||||
Db       5079 GTGGCGAAATTTATTCATGACGATCATCTGCGCCAGGCTTTTTCCTTTCACTCGTTGCTG 5138

Qy       4252 GTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGT 4311
              || ||||||||||||||||||||||| |||||||||||||||||||| || ||||||||
Db       5139 GTCGGCGGTAATCCTTTTGCAACGTCTTCGATCTATACCTTAATTCACGCACTGGAGCGC 5198

Qy       4312 GAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGA 4371
              ||||||||||||||||||||||||||||| ||||| ||||||||  | |||||||||||
Db       5199 GAATGGGGCGTGTGGTTTCCGCGCGGCGGTACCGGTGCGCTGGTTGATGGCATGGCGCGG 5258

Qy       4372 CTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACC 4431
              |||||    || |||||||| || ||| | || || |||||||| ||||||||||| |||
Db       5259 CTGTTTCGCGATTTGGGCGGTGAACTGCTGCTCAACGCCGAAGTCAGCCAGCTGGAGACC 5318

Qy       4432 AGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCT 4491
                 || || ||||| ||||| || ||| |  |||  || || || || |  ||||||||
Db       5319 GAGGGTAACCGCATCAGCGGTGTCCAGCTGAAGGATGGGCGCCGTTTTGCCGCCGCCGCC 5378

Qy       4492 GTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTG 4551
              || || || ||||| ||||||||||||||||||||    ||| |  |||| |||||   |
Db       5379 GTTGCGTCAAATGCTGACGTGGTGCATACCTACGATCGCCTGTTAAGCCAGCATCCTGCG 5438

Qy       4552 GCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTC 4611
              ||    ||||| ||| || |||||||||| ||||| ||||||||||||||||||||||||
Db       5439 GCGCGTAAACGCGCGGCAACGCTGAAGCGCAAGCGGATGAGCAACTCGCTGTTTGTACTC 5498

Qy       4612 TATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCG 4671
              |||||||| || |||||  | ||    ||||| ||||||||||| || || |||||:|||
Db       5499 TATTTTGGTCTTAATCATGCCCACCCGCAGCTGGCGCACCACACGGTGTGCTTTGGTCCG 5558

Qy       4672 CGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCA 4731
              || |||||||| |||||||||||||| ||||| |||||||||||||| || ||||| ||
Db       5559 CGCTATCGTGAATTGATCGATGAGATCTTCAATAGCAGCCAGCTGGCGGAAGATTTCTCG 5618

Qy       4732 CTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGC 4791
              || || ||||| |||||||||  |||||||||||||||||||||| | |||||||||||
Db       5619 CTGTATCTGCATGCGCCCTGCTCCAGCGATCCGTCGCTGGCACCGGCGGGCTGCGGCAGT 5678

Qy       4792 TTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAA 4851
              ||||| ||| | |||||||||||||||||||| ||||| |  || |||||||||||||||
Db       5679 TTTTACGTGCTGGCGCCGGTGCCGCATCTCGGTACCGCCGCAATTGACTGGCAACAGGAA 5738

Qy       4852 GGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTA 4911
              || ||||||||||||||||| |||||||||||||||||| ||||||| ||||||||  |
Db       5739 GGGCCGCGCTTGCGCGATCGCATTTTTGCTTATCTGGAGGAGCACTATATGCCGGGTCTG 5798

Qy       4912 CGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCAT 4971
              || ||||| |||||||||||| | |||||||||||||||||||||||||||||||||||
Db       5799 CGACAGCAGTTAGTGACACACCGTATGTTTACGCCGTTTGATTTTCGCGACACGCTGCAC 5858

Qy       4972 GCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGC 5031
              || ||||| ||||| ||||||||||| || || ||||||||||||||||||||||||||
Db       5859 GCGCATCAGGGCTCAGCGTTTTCGCTCGAACCCATTTTGACGCAAAGCGCCTGGTTCCGG 5918

Qy       5032 CCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCA 5091
              |||||||||||||||||||| || |  || || ||||||||||| || || ||||||||
Db       5919 CCGCATAACCGCGATGCCGACATTACTAACCTTTATCTGGTGGGGGCTGGCACGCATCCC 5978

Qy       5092 GGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAG 5151
              || || || ||||| ||||||||||| || || || || ||||||  ||||||| |||||
Db       5979 GGTGCCGGTGTGCCAGGCGTGATCGGCTCCGCGAAAGCGACCGCCCAGCTGATGGTGGAG 6038

Qy       5152 GATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGG 5211
              ||||   ||| ||||| | || |||  |  ||||||| || |||||||||||||||  ||
Db       6039 GATCTGACCGGATGAACCAACCGCCGCTGATTGAGCAGGTCACGCAAACCATGGCGCAGG 6098

Qy       5212 GCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGC 5271
              |||| || |||||||||| |||  ||  ||| |||||| |  | ||||||||||| ||||
Db       6099 GCTCCAAAAGTTTCGCCAGCGCTACCCGGCTATTTGATCCTTCAACGCGCCGCAGTACGC 6158

Qy       5272 TGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCG 5331
              |||||||||| || |||||||||||||| || |||||||| ||||| || ||||||||||
Db       6159 TGATGCTGTACGCCTGGTGTCGTCACTGTGACGATGTGATAGATGGTCAGACGCTGGGCG 6218

Qy       5332 AAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTG 5391
              |||||||||||||||| || || |  || || ||||| || ||||  || |||||||| |
Db       6219 AAGGCGGCACGCAGCACGCGGTGGCGGATGCACAGGCGCGGATGCGCCACCTGCAAATCG 6278

Qy       5392 AAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGG 5451
              |||||||||||||||||||||| || |||||||||||||| |||||| | || |||||||
Db       6279 AAACCCGCCGCGCCTACAGCGGTGCCCACATGGATGAACCAGCGTTTCGTGCCTTTCAGG 6338

Qy       5452 AAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCG 5511
              |||||||| | |  || ||||| || || || ||||| ||||||||||||||||| || |
Db       6339 AAGTGGCGCTGACGCATCAGCTTCCCCAGCAGCTGGCTTTTGATCATCTGGAAGGGTTTG 6398

Qy       5512 CTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATC 5571
              | |||||||| ||  | |||| ||| ||| | ||||  ||||||||||||||||||||||
Db       6399 CGATGGATGCGCGTGAAGAACGTTATGCGTGTTTCGGGGACACGCTGCGTTACTGCTATC 6458

Qy       5572 ACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGG 5631
              |||| || |||||||| || || |||||||||||||| |||||||| || || ||    |
Db       6459 ACGTGGCCGGCGTGGTGGGGTTAATGATGGCGCGCGTGATGGGCGTACGTGATGAGCGCG 6518

Qy       5632 TGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACA 5691
              | |||||||||||||| ||||| || |||||||| ||||| || || || || || || |
Db       6519 TACTCGATCACGCCTGTGATTTGGGTCTGGCGTTTCAGCTTACCAATATCGCACGGGATA 6578

Qy       5692 TTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGG 5751
              | || || || || ||||| || ||||||||||| |||  |||||| ||| |||| |
Db       6579 TCGTTGAGGACGCGGAGAATGGCCGTTGCTATCTGCCACAAAGCTGGCTGGATGAGGCCG 6638

Qy       5752 GATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAG 5811
              || |  ||||||   ||||  | |  ||||||||||| |||||||| ||| | |||||||
Db       6639 GACTGAGCGCCGCCCAGCTTGCCGATCCGCAACATCGCGCAGCGCTGGCCCCGCTGGCAG 6698

Qy       5812 CGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTT 5871
              ||||| | |||   ||||| || || ||||| || |||||   |||  | || || |
Db       6699 CGCGTCTGGTGCGCGAGGCCGAGCCGTACTATCAGTCAGCGCGCAGCGGGCTGCCAGGAT 6758

Qy       5872 TACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCG 5931
              | ||||| || ||||||||||| ||||| || || ||||||||||| || ||||||||||
Db       6759 TGCCGCTCCGTTCGGCGTGGGCGATCGCCACCGCGCGCGGCGTTTACCGGGAAATTGGCG 6818

Qy       5932 TCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTG 5991
              | ||||| ||||| |||||||  |  || |||||| | || ||||||||||||||||| |
Db       6819 TAAAAGTGCAGCATGCCGGTGCCCGGGCATGGGATACGCGCCAGCGCACCAGTAAAGGCG 6878

Qy       5992 AAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTC 6051
```

```
                |||| |||||||||||||||||||| || ||  | ||| | ||||||||  |  ||| ||
Db         6879 AAAAGCTGGCGCTGCTGGTGAAAGGTGCCGGCGTCGCGCTTACTTCGCGCCTTGCTCATC 6938

Qy         6052 CTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTG 6111
                | ||  ||||||| || ||||||||||||||||| ||||||    ||| || || |||||
Db         6939 CCGAGGCGCGTCCTGCCGGTCTGTGGCAGCGTCCGCGTTGACACGACGCCCATGGCGCTG 6998

Qy         6112 GCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCA 6171
                ||||||||| || ||||||||| | || ||||||| ||||| || || ||||| || |||||
Db         6999 GCGCAGCGTCGCCTGCAGCTTGTGCAACGGTGGGGCGTAAAGAAAGCCAAAGGAAACGCA 7058

Qy         6172 GCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATA 6231
                |||||| || || |||||||||||||||||||||||||||| || || ||||| |||||
Db         7059 GCCTTCCCGTCCCCGCACCGCATGATGCATGCGGTGCGCCATATACAACCGCTTCAGATA 7118

Qy         6232 GCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCAT 6291
                ||||||||| || ||||||||| ||||||||||||||||| |||||||| || ||||||||
Db         7119 GCCTTTGCGTGGAATATAGCGAAACGGCCAGCGTTGATGGACCAGGCCGTCATGCACCAT 7178

Qy         6292 GAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCC 6351
                 || ||||| ||||| || |||||||| || || |||||||||||||||||||| | |||
Db         7179 AAAATAGAGAGCGCCATAGGTCGTCATGCCAGCGCCAATCCACTGCAGCGGCCAGACGCC 7238

Qy         6352 TTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATC 6411
                 |  ||||| |   |||||||  ||||||||  | |||||||||| ||||||| || ||
Db         7239 GTTGACACCCAGCCAAATCAGGCCAATCGACAACAGCGCAAACACCACGGCATACAGGTC 7298

Qy         6412 GTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCA 6471
                |||||||||||| || |     || ||||||||| ||||||| ||||||||||||||||||
Db         7299 GTTGAGCTCAAATTTGCTCTCATGTGGTTCATGATGCGACAAATGCCAGCCCCATCCCCA 7358

Qy         6472 ACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGC 6531
                 || ||||| |||||||||||||||||||||||| ||||| |||||||  | |||||| ||
Db         7359 GCCATGCATAATGTATTTATGCGACAGCGCCGCGACGATCTCCATCAGTATCACGGTAGC 7418

Qy         6532 CAACAAGATAAGCACGTTCCATAACCAGAGCATTGTTCGTCCATTTGTGGAAAAGGGAAG 6591
                || || ||||||| | ||||||||||||||||||  || ||||||||| | | |  |  ||
Db         7419 CAGCAGGATAAGCGCATTCCATAACCAGAGCATCATTGGTCCATTTGCGAAGAGTGAGAG 7478

Qy         6592 TACTAAAGGTGGACGCGGATGAGTGATGGCGCAAG 6626
                || |||||||||||| ||||     | ||||||||
Db         7479 TA-TAAAGGTGGACGTGGATAGCGAAAGGCGCAAG 7512

RESULT 3
US-10-735-019-28
; Sequence 28, Application US/10735019
; Publication No. US20040191863A1
; GENERAL INFORMATION:
;  APPLICANT: E. I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Suh, Wonchul
;  TITLE OF INVENTION: Mutations Affecting Plasmid Copy Number
;  FILE REFERENCE: CL2029 US NA
;  CURRENT APPLICATION NUMBER: US/10/735,019
;  CURRENT FILING DATE:  2003-12-12
;  PRIOR APPLICATION NUMBER: US 60/434973
;  PRIOR FILING DATE: 2002-12-20
;  NUMBER OF SEQ ID NOS: 28
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 28
;   LENGTH: 8609
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid pPCB15
US-10-735-019-28

  Query Match             37.4%;  Score 2619.6;  DB 18;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy          373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
                ||| ||| |     | |  |    |  ||||| || |||||||| |         |  | 
Db         1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy          433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
                ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||| 
Db         1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy          493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
                 |  | ||| |  |  ||||||||||| | || || | | | |  ||    ||  | || 
Db         1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy          553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
                ||| | ||||| || || ||||||||||| || || |||||||||||||||||| || |||
Db         1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy          613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
                |  |||||  | ||||||||||||||| || || || ||| | ||     | |  ||||||
Db         1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy          673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
                 | ||||||||||| || ||  ||||  | || |||   || ||||  ||||||||||   
Db         1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy          733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
                ||    ||| || |  | ||   ||| ||   ||| |  | || || ||| ||||  | 
Db         1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy          793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
                ||| ||||||||  | || || || || |||| ||    || |||    |   ||||
Db         2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy          853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
                ||| | || || ||     | |||||  |  | |||||||||||  |||||||| ||||  | 
Db         2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy          913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
                  |||||||||| | ||||| ||  |||| ||     | | |||  | |  ||       
Db         2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy          973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
                ||| | |  ||| | ||||||||| ||||| |||| ||||| ||  ||||  ||    |  
Db         2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy         1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
                 | ||||| |||||  |  | ||||||   || ||||| ||||||||  |  |  | || 
Db         2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy         1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
                 | | ||  |  |||   ||||||||  | ||  |||||   ||   || ||||||| ||
Db         2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406

Qy         1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
                   || |||||   ||||    ||||   ||||   | ||| || ||   |||| |  ||
Db         2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy         1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
                 | || ||     |||  | ||   |   ||     || ||| ||||||||||||||| |
Db         2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy         1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
                ||||  | | ||||||  ||| |||  ||||     | || ||     |  |||| ||||
Db         2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy         1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
                | |||||  | || ||  | ||||||   ||   |    |  || | |  | |      |
Db         2582 GTCATCGTGTTACGTTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641

Qy         1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
                |||  ||     ||||||| |  || |||||||| ||||||||  |  ||| | || |||
Db         2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy         1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
                | | |   ||    ||   ||| |  ||| |||| ||  ||||| | ||  | ||
Db   2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy   1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
Db   2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy   1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
Db   2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy   1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
Db   2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy   1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
Db   2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy   1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
Db   3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061

Qy   1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
Db   3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy   1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
Db   3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy   1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
Db   3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTCCGTCCCCGGACA 3241

Qy   1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
Db   3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy   2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
Db   3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy   2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
Db   3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy   2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
Db   3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGGATGAATA 3481

Qy   2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
Db   3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy   2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
Db   3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy   2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
Db   3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy   2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
Db   3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy   2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
Db   3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy   2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
Db   3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy   2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
Db   3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy   2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
Db   3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy   2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
Db   3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy   2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
Db   4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy   2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
Db   4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy   2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
Db   4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy   2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
Db   4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy   3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
Db   4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy   3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
Db   4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy   3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
Db   4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy   3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
Db   4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy   3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
Db   4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy   3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
Db   4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy   3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
Db   4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy   3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
Db   4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy   3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
Db   4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy   3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
Db   4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy   3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
Db   4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917
```

```
Qy       3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
Db       4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977
Qy       3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
Db       4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037
Qy       3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
Db       5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097
Qy       3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
Db       5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157
Qy       3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
Db       5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217
Qy       3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
Db       5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277
Qy       4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
Db       5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337
Qy       4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
Db       5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397
Qy       4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
Db       5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457
Qy       4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
Db       5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGGAATCCG 5517
Qy       4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
Db       5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577
Qy       4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
Db       5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637
Qy       4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
Db       5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697
Qy       4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
Db       5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757
Qy       4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
Db       5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817
Qy       4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
Db       5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877
Qy       4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
Db       5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937
Qy       4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
Db       5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997
Qy       4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
```

```
Db       5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057
Qy       4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
Db       6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117
Qy       4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
Db       6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177
Qy       4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
Db       6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237
Qy       4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
Db       6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297
Qy       5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
Db       6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357
Qy       5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
Db       6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417
Qy       5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
Db       6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474
Qy       5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
Db       6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534
Qy       5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
Db       6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594
Qy       5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
Db       6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654
Qy       5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
Db       6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714
Qy       5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
Db       6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774
Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
Db       6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834
Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
Db       6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894
Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
Db       6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954
Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
Db       6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014
Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074
Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
```

```
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | ||||||||||| || ||       || || ||  |||||||||| |||||| | || |
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              || |||  |   | |    |||  | ||||| ||  ||        |   || ||||| |
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
              ||  ||| |||||||| ||        | ||    | || |  |  | |||||| || ||
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
               || |||   | |||||||||||| || |||||||| ||||||||||| ||    || |
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy       6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
               || ||||||||||| ||||| ||||||||||| || || || |||||||||||||| ||
Db       7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy       6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
               ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db       7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy       6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
               || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db       7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy       6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
              ||||||  |||||| ||  |  || || ||||| ||||| |||||||| |  |||||
Db       7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy       6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
               ||  |  |||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db       7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy       6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
              |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db       7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy       6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
               ||||| | ||| | |||  ||  || |
Db       7789 ATTCCAAATCCACAACATAATTTCTCAA 7816
```

```
RESULT 4
US-10-734-778-40
; Sequence 40, Application US/10734778
; Publication No. US20040209365A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Suh, Wonchul
;  APPLICANT:  Rouviere, Pierre
;  TITLE OF INVENTION: PARALLEL CHROMOSOMAL STACKING OF TRAITS IN BACTERIA
;  FILE REFERENCE: CL2026 US NA
;  CURRENT APPLICATION NUMBER: US/10/734,778
;  CURRENT FILING DATE:  2003-12-12
;  PRIOR APPLICATION NUMBER: US 60/434773
;  PRIOR FILING DATE: 2002-12-19
;  NUMBER OF SEQ ID NOS: 43
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
;   LENGTH: 8609
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid pPCB15
```

```
US-10-734-778-40

  Query Match             37.4%;  Score 2619.6;  DB 18;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy        373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
              ||| ||| |     | |  |    |  ||||| || |||||||| |         |  |
Db       1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy        433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
              ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||
Db       1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy        493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
               |  | ||| |  |  ||||||||||| | || || | | | |  ||    ||  | ||
Db       1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy        553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
              ||| | ||||| || || ||||||||||| || || ||||||||||||||||| || |||
Db       1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy        613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
              |  |||||  | |||||||||||||| || || || ||| |  |   | | |  ||||||
Db       1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy        673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
               | ||||||||||| || ||  ||||  | || |||   || ||||  |||||||||
Db       1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy        733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
              ||    ||| || |  | |||   ||| ||   ||| |  | || || ||| ||||  |
Db       1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy        793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
              ||| |||||||  | || || || || ||  |||| ||    || |||    |   ||||
Db       2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy        853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
              ||| | || || ||     | |||||  |  | |||||||||  ||||||| ||||  |
Db       2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
                |||||||||| | ||||| ||  |||| ||       | |||  | |  ||
Db       2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
              ||| | |  ||| | |||||||| |||||  |||| ||||| ||  |||| ||    |
Db       2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
               | ||||| |||||  |  | ||||||   || ||||| ||||||||  |  |  | ||
Db       2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
               | | ||  |  |||   ||||||||  | ||  |||||   || |  | ||||||| ||
Db       2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406

Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
                 || |||||   ||||    ||||   ||||   | ||| || ||   |||| |  ||
Db       2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
               | || ||     |||  | ||   |   ||     || |||| |||||||||||||| |
Db       2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
              ||||  | | ||||||  ||| |||| ||||     | || ||     |  |||| ||||
Db       2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
              | |||||  | || ||  | ||||||   ||   |    |  || | | |  | |      |
Db       2582 GTCATCGTGTTACGTTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641
```

```
Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061

Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
Db       3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
Db       3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
Db       3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTTCCGTCCCCGGACA 3241

Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
Db       3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
Db       3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
Db       3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
Db       3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGGATGAATA 3481

Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
Db       3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
Db       3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy       2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
Db       3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy       2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
Db       3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721
```

```
Qy       2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
Db       3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy       2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
Db       3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy       2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
Db       3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy       2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
Db       3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy       2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
Db       3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy       2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
Db       4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy       2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
Db       4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy       2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
Db       4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy       2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
Db       4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy       3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
Db       4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy       3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
Db       4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy       3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
Db       4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy       3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
Db       4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy       3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
Db       4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy       3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
Db       4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy       3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
Db       4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy       3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
Db       4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy       3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
Db       4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy       3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
```

```
           |  |  || |||  ||||||||| || ||  ||||   | ||  ||||      || ||
Db   4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy   3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
          |||| ||||||||||||| || ||| |    |  ||  |||  ||  |  |||   ||
Db   4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy   3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
                | |   ||  | |||||||  |||   || ||||| || ||||||| ||||||
Db   4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy   3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
          || ||||| |||||  | || || || || || ||        | || ||||||||||||
Db   4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy   3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
          || ||||| || || || ||||| | | ||||    ||||||||| || ||||| || ||
Db   5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy   3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
          || || |||||||||||||||| || |||||  ||||  | ||||| || |||  |||
Db   5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy   3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
          |  ||||||||||||||| |||||||| ||||| || ||||||||||||||||||    |||
Db   5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy   3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
          ||     |  ||||||| ||  | ||| | | | | ||   ||||||      ||| |||
Db   5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy   4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
          ||||  ||||| |  || ||||||    || |||| ||||||| ||||  ||||| |  |||
Db   5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy   4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
          || |||||||| |||||||| ||||| ||| |   | |    |||||||| || | ||||
Db   5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy   4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
          || ||||||| |   ||||| |||||||||||||| |||||| |  || ||    |  |||
Db   5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy   4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
           |||| || ||||| || ||||||||||| || |||||  |  | ||||| || |||||
Db   5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGGAATCCG 5517

Qy   4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
          |||||||| || || || |||  | ||||| ||| | || || ||||||||||| |||
Db   5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy   4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
          ||||| ||||| || ||||| ||||||||   |  |||||      |||||  ||||  ||
Db   5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy   4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
          ||||||||  |  | || || |||   || || ||  ||||||||   ||  |    |||
Db   5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy   4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
             | ||| ||||| || | |||  |||| || ||  ||  ||| ||||| || || ||
Db   5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757

Qy   4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
          || || || ||||||||   | | |||||    || |||||    ||     || |  |||
Db   5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy   4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
          | |   ||| |    ||||| ||||| ||||| |||||||||||||||||||||| ||  ||
Db   5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy   4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
          || |  || || || ||||| || || |||||||||||||| || || || |||||  ||
```

```
Db   5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy   4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
          ||  | || ||||| |||         |||| || |||||||| |||||   | ||||| 
Db   5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy   4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
          || ||   ||  |||||||| |||||||||   || |||||||||| ||||||| | |||
Db   5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy   4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
          || || || ||  | ||||| ||  || ||||||||||      |||||||| ||  |||||
Db   6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy   4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
          ||||| |||||||  || || ||||| || |||||||| || || |     || || |||
Db   6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy   4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
          || ||| | |||||||||||||| ||||| ||||||  |||  |||||   || || |||
Db   6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy   4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
          || || |||  | || || || |||| || |||||||||||||||| || ||||||||||||
Db   6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy   5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
              | ||    ||||| |||||||| || || || || || ||||| ||||||||| | |||
Db   6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy   5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
          ||||| ||||| ||||| ||||| || || ||  |  | ||||||||||| |       | ||
Db   6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy   5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
          ||     | | |||||  | || |      ||||||||||||| |||||||||| || || |
Db   6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy   5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
          | || ||  | | ||| || || ||    || || |||||  |||||||||| || || |
Db   6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy   5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
          |||| || ||||||||| || || ||||| |  ||||| ||||||      | |   ||||
Db   6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy   5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
             |     |     |  ||   ||  ||||||| ||  ||||  ||||  || |  ||||
Db   6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy   5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
          ||  |||  ||||| ||| | || || || |||   ||||||||||| || |||||    |
Db   6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy   5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
          |  |  |  | |        ||||| || |||||||||||| || || ||||||  |||
Db   6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774

Qy   5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
          |    |  |||  ||   | || || |||||||||| |||||||||||||| || || |
Db   6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy   5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
          | ||  ||||||||||||   | |||||||| |||||  | ||   ||||||||| |||||
Db   6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy   5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
          ||||| | || ||||| |||||||| | || |||||||||||| || ||||| || |||||
Db   6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy   5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
          |    || |||| ||||||||  ||  ||||||| ||  |||  ||| |       |  |
Db   6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014
```

```
Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                    |||| ||  || | ||    ||  |   |    | || | |||  | ||   ||
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
              | ||||||||||||||    ||  |     ||||| |  |    |||||  | |||||||
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | ||||||||||| || ||       || || ||  |||||||||| |||||| | || |
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              || |||  |   | |||  |||  | ||||| ||  ||        |  ||| ||||| |
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
              ||  ||| |||||||| ||        | ||    | || |  |  | |||||| || ||
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
               || |||   | |||||||||||| || |||||||| ||||||||||| ||    || |
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy       6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
               || ||||||||||| ||||| ||||||||||| || || || |||||||||||||| ||
Db       7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy       6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
               ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db       7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy       6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
               || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db       7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy       6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
              ||||||  |||||| ||  |  || || ||||| ||||| |||||||| |  |||||
Db       7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy       6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
               ||  |  |||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db       7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy       6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
              |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db       7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy       6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
               ||||| | ||| | |||  ||  || |
Db       7789 ATTCCAAATCCACAACATAATTTCTCAA 7816
```

```
RESULT 5
US-10-734-936-37
; Sequence 37, Application US/10734936
; Publication No. US20040209370A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Co., Inc.
;  APPLICANT:  Suh, Wonchul
;  TITLE OF INVENTION: Method for Chromosomal Engineering
;  FILE REFERENCE: CL1878 US NA
;  CURRENT APPLICATION NUMBER: US/10/734,936
;  CURRENT FILING DATE:  2003-12-12
;  PRIOR APPLICATION NUMBER: US 60/434602
;  PRIOR FILING DATE: 2002-12-19
;  NUMBER OF SEQ ID NOS: 37
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
;  LENGTH: 8609
;  TYPE: DNA
;  ORGANISM: Artificial sequence
;  FEATURE:
;  OTHER INFORMATION: Plasmid pPCB15
US-10-734-936-37

  Query Match             37.4%;  Score 2619.6;  DB 18;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;
```

```
Qy        373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
              ||| ||| |     | |  |    |  ||||| || |||||||| |         |  |
Db       1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy        433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
              ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||
Db       1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy        493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
               |  | ||| |  |  ||||||||||| | || || | | | |  ||    ||  | ||
Db       1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy        553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
              ||| | ||||| || || ||||||||||| || || ||||||||||||||||| || |||
Db       1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy        613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
              |  |||||  | |||||||||||||| || || || ||| | ||     | |  ||||||
Db       1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy        673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
               | ||||||||||| || ||  ||||  | || |||   || ||||  |||||||||
Db       1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy        733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
              ||    ||| || |  | |||   ||| ||   ||| |  | || || ||| ||||  |
Db       1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy        793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
              ||| |||||||  | || || || || ||  |||| ||    || |||    |   ||||
Db       2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy        853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
              ||| | || || ||     | |||||  |  | |||||||||  ||||||| ||||  |
Db       2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
                |||||||||| | ||||| ||  |||| ||     | | |||  | |  ||
Db       2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
              ||| | |  ||| | |||||||| |||||  |||| ||||| ||  |||| ||    |
Db       2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
               | ||||| |||||  |  | ||||||   || ||||| ||||||||  |  |  | ||
Db       2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
               | | ||  |  |||   ||||||||  | ||  |||||   ||   || ||||||| ||
Db       2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406

Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
                 || |||||   ||||    ||||   ||||   | ||| || ||   |||| |  ||
Db       2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
               | || ||     |||  | ||   |   ||     || |||| |||||||||||||| |
Db       2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
```

```
Db       2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
Db       2582 GTCATCGTGTTACGTTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641

Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061

Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
Db       3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
Db       3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
Db       3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTCCGTCCCCGGACA 3241

Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
Db       3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
Db       3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
Db       3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
Db       3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGATGAATA 3481

Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
Db       3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
Db       3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy       2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
Db       3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy       2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
Db       3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy       2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
Db       3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy       2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
Db       3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy       2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
Db       3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy       2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
Db       3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy       2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
Db       3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy       2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
Db       4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy       2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
Db       4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy       2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
Db       4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy       2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
Db       4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy       3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
Db       4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy       3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
Db       4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy       3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
Db       4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy       3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
Db       4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy       3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
Db       4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy       3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
Db       4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy       3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
Db       4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy       3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
Db       4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737
```

```
Qy   3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
          |||||| |||||   | |  ||   |||||||| |||||||| ||||| ||  |  |
Db   4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy   3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
           |  |  || |||  |||||||||| || ||  ||||   | ||  ||||      || ||
Db   4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy   3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
          |||| ||||||||||||| || ||| |    |  ||  |||  ||  |  |||   ||
Db   4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy   3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
                | |   ||  | ||||||  |||   || ||||| || |||||||| ||||||
Db   4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy   3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
          || ||||| |||||  | || || || || ||     | || ||||||||||||
Db   4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy   3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
          || ||||| || || || ||||| | | ||||    ||||||||| || ||||| || ||
Db   5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy   3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
          || || |||||||||||||||| || |||||  ||||  | ||||| || |||  |||
Db   5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy   3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
          |  ||||||||||||||| ||||||| ||||| || |||||||||||||||||    |||
Db   5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy   3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
          ||     |  ||||||| ||  | ||| | | | | ||   ||||||      ||| |||
Db   5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy   4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
          ||||  ||||| |  || ||||    || |||| ||||| ||||  ||||| |  |||
Db   5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy   4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
          || ||||||||| ||||||||| ||||| ||| |   | |    |||||||| || | ||||
Db   5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy   4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
          || ||||||| |   ||||| |||||||||||||| |||| | |  || ||    |  |||
Db   5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy   4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCT 4266
           |||| || ||||| || ||||||||||| || |||||  |  | ||||| || |||||
Db   5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGGAATCCG 5517

Qy   4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
          |||||||| || || || |||||  | ||||| ||| | || || ||||||||||| |||
Db   5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy   4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
          ||||| ||||| || ||||| ||||||||   | ||||||      |||||  ||||  ||
Db   5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy   4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
          ||||||||  |  | || || |||   || || ||  |||||||||   ||  |    |||
Db   5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy   4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
             | ||| ||||| || | |||  |||| || ||  ||  ||| ||||| || || ||
Db   5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757

Qy   4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
          || || || ||||||||   | | |||||    || |||||    ||    || |  |||
Db   5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy   4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
          | |   ||| |    ||||| ||||| ||||| |||||||||||||||||||| || ||
Db   5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy   4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
          || |  || || || ||||| || || |||||||||||||| || || || |||||  ||
Db   5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy   4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
          ||  | || ||||| |||         ||||| || |||||||| |||||  | |||||
Db   5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy   4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
          || ||   ||  |||||||| |||||||||   || |||||||||| ||||||| | |||
Db   5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy   4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
          || || || ||  | ||||| ||  || ||||||||      |||||||| ||  |||||
Db   6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy   4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
          ||||| |||||||  || || ||||| || |||||||| || || ||    || || |||
Db   6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy   4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
          || ||| | |||||||||||||| ||||| ||||||  |||  |||||   || || |||
Db   6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy   4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
          || || ||| | || || ||| |||| || |||||||||||||| || ||||||||||||
Db   6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy   5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
              | ||    ||||| |||||||| || || || || || |||| ||||||||| | |||
Db   6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy   5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
          ||||| ||||| ||||| ||||| ||  |  | ||||||||||| |       | ||
Db   6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy   5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
          ||     | | |||||  | || |      ||||||||||||| |||||||| || || |
Db   6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy   5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
          | || ||  | | ||| || || ||    || || |||||  ||||||||| || || |
Db   6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy   5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
          |||| || |||||||| || || ||||| |  ||||| ||||||      | |   ||||
Db   6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy   5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
             |     |     |  ||   ||  |||||||| ||  ||||||  |||| || |  ||||
Db   6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy   5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
          ||  |||  |||| ||| | || || || |||   |||||||||| || |||||    |
Db   6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy   5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
          |  |  |  | |        ||||| || ||||||||||| || || |||||||  |||
Db   6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774

Qy   5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
          |    |  |||  ||   | || || ||||||||||| |||||||||||||| || || |
Db   6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy   5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
          | ||  ||||||||||||   | ||||||||| |||||  | ||   |||||| |||||
Db   6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy   5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
```

```
              ||||| | || ||||| |||||| | || ||||||||||| || ||||| || |||||
Db       6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
              |    || ||||| ||||||||  ||  |||||| ||  |||  ||| |       |  |
Db       6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                    |||| ||  || | ||    ||  |   |    | || | |||  | ||   ||
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
              | ||||||||||||||    ||  |     ||||| |  |    |||||  | |||||||
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | ||||||||||| || ||       || || ||  |||||||||| |||||| | || |
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              || |||  |   | |||  |||  | ||||| ||  ||        |  ||| ||||| |
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
              ||  ||| ||||||||| ||        | ||    | || |  |  | ||||||| || ||
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
               || |||   | |||||||||||||| ||||||||| ||||||||||||| ||    || |
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy       6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
               || |||||||||||| ||||| |||||||||||| || || |||||||||||||||| ||
Db       7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy       6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
               ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db       7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy       6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
               || |  ||||| || |||||||||||||| |||||||| || |||        ||||  ||
Db       7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy       6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
              ||||||  ||||||| ||  |  || || |||| ||||| |||||||| |  |||||
Db       7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy       6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
               ||  |  |||||||||| || || |||||||||||||| ||||| ||||||||||||||
Db       7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy       6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
              |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db       7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy       6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
               ||||| | ||| | |||  ||  || |
Db       7789 ATTCCAAATCCACAACATAATTTCTCAA 7816

RESULT 6
US-10-735-442-64
; Sequence 64, Application US/10735442
; Publication No. US20040219629A1
; GENERAL INFORMATION:
;  APPLICANT: E. I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Suh, Wonchul
;  APPLICANT:  Rouviere, Pierre
;  APPLICANT:  Cheng, Qiong
;  TITLE OF INVENTION: Increasing Carotenoid Production in Bacteria Via Chromosomal
;  TITLE OF INVENTION:  Integration
;  FILE REFERENCE: CL2027 US NA
;  CURRENT APPLICATION NUMBER: US/10/735,442
;  CURRENT FILING DATE:  2003-12-12
;  PRIOR APPLICATION NUMBER: US 60/434618
;  PRIOR FILING DATE: 2002-12-19
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 64
;   LENGTH: 8609
;   TYPE: DNA
;   ORGANISM: Artificial sequence
;   FEATURE:
;   OTHER INFORMATION: Plasmid pPCB15
US-10-735-442-64

  Query Match             37.4%;  Score 2619.6;  DB 18;  Length 8609;
  Best Local Similarity   64.7%;  Pred. No. 0;
  Matches 4015;  Conservative    0;  Mismatches 2169;  Indels   24;  Gaps    7;

Qy        373 CAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCAAAGC 432
              ||| ||| |     | |  |    |  ||||| || ||||||||| |         |  |
Db       1627 CAGTTGCTGGCTGATATCGATAGCCGCCTTGATCAGTTACTGCCGGTTCAGGGTGAGCGG 1686

Qy        433 GATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCGTCCT 492
              ||| | ||| |||||||||||||||  || |||||||| | ||||||||||||||||||
Db       1687 GATTGTGTGGGTGCCGCGATGCGTGAAGGCACGCTGGCACCGGGCAAACGTATTCGTCCG 1746

Qy        493 TTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGTTCTC 552
               |  | ||| |  |  ||||||||||| | || || | | | |  ||     ||  | ||
Db       1747 ATGCTGCTGTTATTAACAGCGCGCGATCTTGGCTGTGCGATCAGTCACGGGGGATTACTG 1806

Qy        553 GATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCC 612
              ||| | ||||| || || ||||||||| || || ||||||||||||||||||| || |||
Db       1807 GATTTAGCCTGCGCGGTTGAAATGGTGCATGCTGCCTCGCTGATTCTGGATGATATGCCC 1866

Qy        613 TCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAA 672
              |  |||||  | |||||||||||||| || || || ||| | ||     | |  ||||||
Db       1867 TGCATGGACGATGCGCAGATGCGTCGGGGGCGTCCCACCATTCACACGCAGTACGGTGAA 1926

Qy        673 AACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATT 732
               | ||||||||||| || ||  ||||  | || |||   || ||||  |||||||||
Db       1927 CATGTGGCGATTCTGGCGGCGGTCGCTTTACTCAGCAAAGCGTTTGGGGTGATTGCCGAG 1986

Qy        733 GCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTC 792
              ||    ||| || |  | |||   ||| ||   ||| |  | || || || ||||  |
Db       1987 GCTGAAGGTCTGACGCCGATAGCCAAAACTCGCGCGGTGTCGGAGCTGTCCACTGCGATT 2046

Qy        793 GGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGC 852
              ||| |||||||  | || || || || ||  |||| ||    || |||    |   ||||
Db       2047 GGCATGCAGGGTCTGGTTCAGGGCCAGTTTAAGGACCTCTCGGAAGGCGATAAACCCCGC 2106

Qy        853 AGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACG 912
              ||| | || || ||     | |||||  |  | |||||||||  |||||||| ||||  |
Db       2107 AGCGCCGATGCCATACTGCTAACCAATCAGTTTAAAACCAGCACGCTGTTTTGCGCGTCA 2166

Qy        913 CTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTC 972
                |||||||||| | ||||| ||  ||||||     | | |||  | |  ||
Db       2167 ACGCAAATGGCGTCCATTGCGGCCAACGCGTCCTGCGAAGCGCGTGAGAACCTGCATCGT 2226

Qy        973 TTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAA 1032
              ||| | |  ||| | |||||||| ||||| |||| |||||| ||  |||| ||    |
Db       2227 TTCTCGCTCGATCTCGGCCAGGCCTTTCAGTTGCTTGACGATCTTACCGATGGCATGACC 2286

Qy       1033 CACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGT 1092
               | ||||| |||||  |  |  ||||||   || ||||| |||||||  |  |  | ||
Db       2287 GATACCGGCAAAGACATCAATCAGGATGCAGGTAAATCAACGCTGGTCAATTTATTAGGC 2346

Qy       1093 GCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCTTGCC 1152
               | | ||  |  |||   ||||||||  | ||  |||||   ||   || ||||||| ||
Db       2347 TCAGGCGCGGTCGAAGAACGCCTGCGACAGCATTTGCGCCTGGCCAGTGAACACCTTTCC 2406
```

```
Qy       1153 TGCGCCTGCCATCGCGGCATCGCCACTCGCCAA---TATATGCACGCGCTGTTTAATCAA 1209
                 || |||||   ||||    ||||   ||||   | ||| || ||   |||| |  ||
Db       2407 GCGGCATGCCAAAACGGCCATTCCACCACCCAACTTTTTATTCAGGCCTGGTTTGACAAA 2466

Qy       1210 CAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCGC 1269
               | || ||     |||  | ||   |   ||     || |||| |||||||||||||| |
Db       2467 AAACTCGCTGCCGTCAGTTAAGGATGCTGCA-----TGAGCCATTTTGCGGTGATCGCAC 2521

Qy       1270 CGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGCG 1329
              ||||  | | ||||||  ||| |||  ||||     | || ||     |  |||| ||||
Db       2522 CGCCCTTTTTCAGCCATGTTCGCGCTCTGCAAAACCTTGCTCAGGAATTAGTGGCCCGCG 2581

Qy       1330 GCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGCA 1389
              | |||||  | || ||  | ||||||   ||   |    |  || | |  | |      |
Db       2582 GTCATCGTGTTACGTTTTTTCAGCAACATGACTGCAAAGCGCTGGTAACGGGCAGCGATA 2641

Qy       1390 TCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTGC 1449
              |||  ||     ||||||| |  || ||||||||| |||||||  |  ||| | || |||
Db       2642 TCGGATTCCAGACCGTCGGACTGCAAACGCATCCTCCCGGTTCCTTATCGCACCTGCTGC 2701

Qy       1450 ATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCCT 1509
              | | |   ||    ||   ||| |  ||| |||| ||  ||||| | ||  | ||
Db       2702 ACCTGGCCGCGCACCCACTCGGACCCTCGATGTTACGACTGATCAATGAAATGGCACGTA 2761

Qy       1510 GCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGCG 1569
               || ||||||||| ||||| |||||||| || |   |  | || ||| | || || ||||
Db       2762 CCAGCGATATGCTTTGCCGGGAACTGCCCGCCGCTTTTCATGCGTTGCAGATAGAGGGCG 2821

Qy       1570 TGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCGT 1629
              ||||||  ||  |||||||  |||| || | | | ||||| ||||||     ||||||||
Db       2822 TGATCGTTGATCAAATGGAGCCGGCAGGTGCAGTAGTCGCAGAAGCGTCAGGTCTGCCGT 2881

Qy       1630 TTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTGA 1689
              |||||||||||||||||||  ||||| |||| || ||| | ||  | || || |||||||
Db       2882 TTGTTTCGGTGGCCTGCGCGCTGCCGCTCAACCGCGAACCGGGTTTGCCTCTGGCGGTGA 2941

Qy       1690 TGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGATA 1749
              |||| |||   |  |        ||    || | | |||| | |    |||||    | |
Db       2942 TGCCTTTCGAGTACGGCACCAGCGATGCGGCTCGGGAACGCTATACCACCAGCGAAAAAA 3001

Qy       1750 TCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTTA 1809
              | |||||  |  | ||||| |||||||     ||||||    | || ||  |     |
Db       3002 TTTATGACTGGCTGATGCGACGTCACGATCGTGTGATCGCGCATCATGCATGCAGAATGG 3061

Qy       1810 ATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAGA 1869
               |||  |   |||      | | ||||| ||  | || || ||||||||||||||||||
Db       3062 GTTTAGCCCCGCGTGAAAAACTGCATCATTGTTTTTCTCCACTGGCACAAATCAGCCAGT 3121

Qy       1870 TGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGGC 1929
              || | || |   | |||||||| ||  |    ||||| | ||||| ||| || || || |
Db       3122 TGATCCCCGAACTGGATTTTCCCCGCAAAGCGCTGCCAGACTGCTTTCATGCGGTTGGAC 3181

Qy       1930 CACTCCGCG---CCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTC 1986
              |  | ||     ||| ||   | || | | |      | |    |  ||  | | |
Db       3182 CGTTACGGCAACCCCAGGGGACGCCGGGGTCATCAACTTCTTATTTTCCGTCCCCGGACA 3241

Qy       1987 AGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGC 2046
              | ||     ||| |||||||||||| || ||||| || ||||| |   | |||||   |
Db       3242 AACCCCGTATTTTTGCCTCGCTGGGCACCCTGCAGGGACATCGTTATGGCCTGTTCAGGA 3301

Qy       2047 ATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGAT 2106
                 | ||  | || |||    || ||        | |  || | || || ||||| ||
Db       3302 CCATCGCCAAAGCCTGCGAAGAGGTGGATGCGCAGTTACTGTTGGCACACTGTGGCGGCC 3361

Qy       2107 TAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCG 2166
              |    |||   ||| |   | | ||||  |  |  || |       |||   |||||| |
Db       3362 TCTCAGCCACGCAGGCAGGTGAACTGGCCCGGGGCGGGGACATTCAGGTTGTGGATTTTG 3421

Qy       2167 TCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACA 2226
               ||||||   |||||| ||    || || ||| ||   ||||| ||||  ||| | || |
Db       3422 CCGATCAATCCGCAGCACTTTCACAGGCACAGTTGACAATCACACATGGTGGGATGAATA 3481

Qy       2227 GCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATC 2286
                |  ||||| ||  | |  | | | || ||| | |||||||||||| | || |||||||
Db       3482 CGGTACTGGACGCTATTGCTTCCCGCACACCGCTACTGGCGCTGCCGCTGGCATTTGATC 3541

Qy       2287 AGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTA 2346
              | || ||||||||  | || ||||  |  |||| | | ||    || || || || ||||
Db       3542 AACCTGGCGTGGCATCACGAATTGTTTATCATGGCATCGGCAAGCGTGCGTCTCGGTTTA 3601

Qy       2347 GCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGC 2406
                      |||   |||| || |||  | | |  |||||||||  ||     |||| |||
Db       3602 CTACCAGCCATGCGCTGGCGCGGCAGATTCGATCGCTGCTGACTAACACCGATTACCCGC 3661

Qy       2407 TACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCG 2466
                || ||| ||   ||||||||     |||  |  ||||||||    |     |||||||
Db       3662 AGCGTATGACAAAAATTCAGGCCGCATTGCGTCTGGCAGGCGGCACACCAGCCGCCGCCG 3721

Qy       2467 ACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCG 2526
              | || || || |||||| || |  | | |  || |        |  |||    | ||||
Db       3722 ATATTGTTGAACAGGCGATGCG-GACCTGTCAGCCAGTACTCAGTGGGCAGGATTATGCA 3780

Qy       2527 CACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCG 2586
                |||| || ||| | ||| |||||||||| || ||||| |||||| | || ||||| ||
Db       3781 ACCGCACTATGATCTCATTCTGGTCGGTGCCGGTCTGGCTAATGGCCTTATCGCGCTCCG 3840

Qy       2587 TCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGC 2646
               || |  |||  ||| ||  |  ||     | || |  | |||   | | |  |  |||
Db       3841 GCTTCAGCAACAGCATCCGGATATGCGGATCTTGCTTATTGAGGCGGGTCCTGAGGCGGG 3900

Qy       2647 AGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCT 2706
              ||| || ||||||||||| |||||  |    ||| | |      | || | ||||||| |
Db       3901 AGGGAACCATACCTGGTCCTTTCACGAAGAGGATTTAACGCTGAATCAGCATCGCTGGAT 3960

Qy       2707 GCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCG 2766
                  |||||  |   |   |  ||| | |  || ||||| ||||| ||    |  || ||
Db       3961 AGCGCCGCTTGTGGTCCATCACTGGCCCGACTACCAGGTTCGTTTCCCCCAACGCCGTCG 4020

Qy       2767 CAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGC 2826
              | || || || | |  || |  | | |  | ||  |  |||| ||| |    ||
Db       4021 CCATGTGAACAGTGGCTACTACTGCGTGACCTCCCGGCATTTCGCCGGGATACTCCGGCA 4080

Qy       2827 GGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGT 2886
                 | | ||  |  || | |||    | || || ||   |   ||  |  |      |||
Db       4081 ACAGTTTGGACAACATTTATGGCTGCATACCGCGGTTTCAGCCGTTCATGCTGAATCGGT 4140

Qy       2887 GACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCA 2946
                 | | |||||||||||     ||  |||       ||||| || || || ||
Db       4141 CCAGTTAGCGGATGGCCGGATTATTCATGCCAGTACAGTGATCGACGGACGGGGTTACAC 4200

Qy       2947 GCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGC 3006
              |||     |    || |   | || |  ||||  ||| | || || |||||||| |||
Db       4201 GCCTGATTCTGCACTACGCGTAGGATTCCAGGCATTTATCGGTCAGGAGTGGCAACTGAG 4260

Qy       3007 GCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGG 3066
                 |||||| ||  |   |   |||||  | |||||||| || ||||||||||||   ||
Db       4261 CGCGCCGCATGGTTTATCGTCACCGATTATCATGGATGCGACGGTCGATCAGCAAAATGG 4320

Qy       3067 TTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCA 3126
               || || ||||| || || ||||||||   |||       ||  |||| ||||| || ||
Db       4321 CTACCGCTTTGTTTATACCCTGCCGCTTTCCGCAACCGCACTGCTGATCGAAGACACACA 4380

Qy       3127 TTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGC 3186
               ||| || || || |    ||   ||   |    || |||||| ||||   ||| |||||
Db       4381 CTACATTGACAAGGCTAATCTTCAGGCCGAACGGGCGCGTCAGAACATTCGCGATTATGC 4440

Qy       3187 CAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGAT 3246
                  |  || || ||| || |    ||| ||||||| ||||| || ||   ||| || ||
Db       4441 TGCGCGACAGGGTTGGCCGTTACAGACGTTGCTGCGGGAAGAACAGGGTGCATTGCCCAT 4500

Qy       3247 TACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGG 3306
              |||  | |  ||| | |    ||  || ||||||||||| |     ||||| || |||||
Db       4501 TACGTTAACGGGCGATAATCGTCAGTTTTGGCAACAGCAAC---CGCAAGCCTGTAGCGG 4557

Qy       3307 CCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCT 3366
                | |||||||||||||||||| | || ||||| |||||| | |||  ||| ||||||||
Db          4558 ATTACGCGCCGGGCTGTTTCATCCGACAACCGGCTACTCCCTACCGCTCGCGGTGGCGCT 4617

Qy          3367 AGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGA 3426
                  || ||     |    ||||||             |   | |  |   |||    || |
Db          4618 GGCCGATCGTCTCAGCGCGCTGGATGTGTTTACCTCTTCCTCTGTTCACCAGACGATTGC 4677

Qy          3427 ACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCT 3486
                  | ||||||||  || |  ||||   ||||| | ||||||||  |||| || |||||| |
Db          4678 TCACTTTGCCCAGCAACGTTGGCAGCAACAGGGGTTTTTCCGCATGCTGAATCGCATGTT 4737

Qy          3487 GTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGA 3546
                 |||||| |||||   | |  ||   |||||||| |||||||| ||||| ||  |  |
Db          4738 GTTTTTAGCCGGACCGGCCGAGTCACGCTGGCGTGTGATGCAGCGTTTCTATGGCTTACC 4797

Qy          3547 TGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGAT 3606
                  |  |  || |||  ||||||||| || ||  ||||   | ||  ||||      || ||
Db          4798 CGAGGATTTGATTGCCCGCTTTTATGCGGGAAAACTCACCGTGACCGATCGGCTACGCAT 4857

Qy          3607 TCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGT 3666
                 |||| ||||||||||||| || ||| |    |  ||  |||  ||  |  |||   ||
Db          4858 TCTGAGCGGCAAGCCGCCCGTTCCCGTTTTCGCGGCATTGCAGGCAATTATGACGACTCA 4917

Qy          3667 CGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTG 3726
                       | |   ||  | |||||||  |||   || ||||| || |||||||| ||||||
Db          4918 TCGTTGAAGAGCGACTACATGAAACCAACTACGGTAATTGGTGCGGGCTTTGGTGGCCTG 4977

Qy          3727 GCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGAC 3786
                 || ||||| |||||  | || || || || || ||        | || ||||||||||||
Db          4978 GCACTGGCAATTCGTTTACAGGCCGCAGGTATTCCTGTTTTGCTGCTTGAGCAGCGCGAC 5037

Qy          3787 AAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCC 3846
                 || ||||| || || || ||||| | | ||||    |||||||| || ||||| || ||
Db          5038 AAGCCGGGTGGCCGGGCTTATGTTTATCAGGAGCAGGGCTTTACTTTTGATGCAGGCCCT 5097

Qy          3847 ACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTC 3906
                 || || ||||||||||||||||| || |||||  ||||  | ||||| || |||  |||
Db          5098 ACCGTTATCACCGATCCCAGCGCGATTGAAGAACTGTTTGCTCTGGCCGGTAAACAGCTT 5157

Qy          3907 AGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGC 3966
                 |  ||||||||||||||| ||||||| ||||| || |||||||||||||||||    |||
Db          5158 AAGGATTACGTCGAGCTGTTGCCGGTCACGCCGTTTTATCGCCTGTGCTGGGAGTCCGGC 5217

Qy          3967 AAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAAT 4026
                 ||     |  ||||||| ||  | ||| | | | | ||   ||||||      ||| |||
Db          5218 AAGGTCTTCAATTACGATAACGACCAGGCCCAGTTAGAAGCGCAGATACAGCAGTTTAAT 5277

Qy          4027 CCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAG 4086
                 ||||  ||||| |  || |||||    || |||| |||||| ||||  ||||| |  |||
Db          5278 CCGCGCGATGTTGCGGGTTATCGAGCGTTCCTTGACTATTCGCGTGCCGTATTCAATGAG 5337

Qy          4087 GGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCG 4146
                 || |||||||| |||||||| ||||| ||| |   | |    |||||||| || | ||||
Db          5338 GGCTATCTGAAGCTCGGCACTGTGCCTTTTTTATCGTTCAAAGACATGCTTCGGGCCGCG 5397

Qy          4147 CCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATT 4206
                 || ||||||| |   ||||| |||||||||||||| ||||| |  || ||    |  |||
Db          5398 CCCCAGTTGGCAAAGCTGCAGGCATGGCGCAGCGTTTACAGTAAAGTTGCCGGCTACATT 5457

Qy          4207 CAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGCGGTAATCCT 4266
                  |||| || ||||| || ||||||||||| || |||||  |  | ||||  || |||||
Db          5458 GAGGATGAGCATCTTCGGCAGGCGTTTTCTTTTCACTCGCTCTTAGTGGGGGGAATCCG 5517

Qy          4267 TTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGG 4326
                 |||||||| || || || |||||  | ||||| ||| | || || ||||||||||| |||
Db          5518 TTTGCAACCTCGTCCATTTATACGCTGATTCACGCGTTAGAACGGGAATGGGGCGTCTGG 5577

Qy          4327 TTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTG 4386
                 ||||| ||||| || ||||| ||||||||  | ||||||      |||||  ||||  ||
Db          5578 TTTCCACGCGGTGGAACCGGTGCGCTGGTCAATGGCATGATCAAGCTGTTTCAGGATCTG 5637

Qy          4387 GGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATT 4446
                 ||||||||  |  | || || |||   || || ||  ||||||||   ||  |    |||
Db          5638 GGCGGCGAAGTCGTGCTTAACGCCCGGGTCAGTCATATGGAAACCGTTGGGGACAAGATT 5697

Qy          4447 AGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCC 4506
                    | ||| ||||| || | |||  |||| || ||  ||  ||| ||||| || || ||
Db          5698 CAGGCCGTGCAGTTGGAAGACGGCAGACGGTTTGAAACCTGCGCGGTGGCGTCGAACGCT 5757

Qy          4507 GACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCG 4566
                 || || || ||||||||   | | |||||    || |||||    ||    || |  |||
Db          5758 GATGTTGTACATACCTATCGCGATCTGCTGTCTCAGCATCCCGCAGCCGCTAAGCAGGCG 5817

Qy          4567 ACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAAT 4626
                 | |   ||| |    ||||| ||||| ||||| |||||||||||||||||||| || ||
Db          5818 AAAAAACTGCAATCCAAGCGTATGAGTAACTCACTGTTTGTACTCTATTTTGGTCTCAAC 5877

Qy          4627 CAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTG 4686
                 || |  || || || ||||| || || |||||||||||||| || || || |||||  ||
Db          5878 CATCATCACGATCAACTCGCCCATCATACCGTCTGTTTTGGGCCACGCTACCGTGAACTG 5937

Qy          4687 ATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCG 4746
                 ||  | || ||||| |||         ||||| || |||||||| |||||  | |||||
Db          5938 ATTCACGAAATTTTTAACCATGATGGTCTGGCTGAGGATTTTTCGCTTTATTTACACGCA 5997

Qy          4747 CCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCG 4806
                 || ||   ||  |||||||| |||||||||   || |||||||||| ||||||| | |||
Db          5998 CCTTGTGTCACGGATCCGTCACTGGCACCGGAAGGGTGCGGCAGCTATTATGTGCTGGCG 6057

Qy          4807 CCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGC 4866
                 || || || ||  | ||||| ||  || ||||||||      |||||||| ||  |||||
Db          6058 CCTGTTCCACACTTAGGCACGGCGAACCTCGACTGGGCGGTAGAAGGACCCCGACTGCGC 6117

Qy          4867 GATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTG 4926
                 ||||| |||||||  || || ||||| || |||||||| || || ||    || || |||
Db          6118 GATCGTATTTTTGACTACCTTGAGCAACATTACATGCCTGGCTTGCGAAGCCAGTTGGTG 6177

Qy          4927 ACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCG 4986
                 || ||| | |||||||||||||| ||||| ||||||  |||  |||||   || || |||
Db          6178 ACGCACCGTATGTTTACGCCGTTCGATTTCCGCGACGAGCTCAATGCCTGGCAAGGTTCG 6237

Qy          4987 GCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGAT 5046
                 || || ||| | || || ||| |||| || |||||||||||||| || ||||||||||||
Db          6238 GCCTTCTCGGTTGAACCTATTCTGACCCAGAGCGCCTGGTTCCGACCACATAACCGCGAT 6297

Qy          5047 GCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCC 5106
                     | ||    ||||| |||||||| || || || || ||||| ||||||||| | |||
Db          6298 AAGCACATTGATAATCTTTATCTGGTTGGCGCAGGCACCCATCCTGGCGCGGGCATTCCC 6357

Qy          5107 GGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
                 ||||| ||||| ||||| ||||| || ||  |  | ||||||||||| |       | ||
Db          6358 GGCGTAATCGGCTCGGCGAAGGCGACGGCAGGCTTAATGCTGGAGGACCTGATTTGACGA 6417

Qy          5167 ATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCG 5226
                 ||     | | |||||  | || |      ||||||||||||| |||||||| || || |
Db          6418 AT---ACGTCATTACTGAATCATGCCGTCGAAACCATGGCGGTTGGCTCGAAAAGCTTTG 6474

Qy          5227 CCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGT 5286
                 | || ||  | | ||| || || ||    || || ||||||  ||||||||| || || |
Db          6475 CGACTGCATCGACGCTTTTCGACGCCAAAACCCGTCGCAGCGTGCTGATGCTTTACGCAT 6534

Qy          5287 GGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGC 5346
                 |||| || |||||||| || || ||||| |  ||||| ||||||      | |   ||||
Db          6535 GGTGCCGCCACTGCGACGACGTCATTGACGATCAAACACTGGGCTTTCATGCCGACCAGC 6594

Qy          5347 ATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCT 5406
                    |     |     |  ||   ||  ||||||| || |||||  |||| || |  ||||
Db          6595 CCTCTTCGCAGATGCCTGAGCAGCGCCTGCAGCAGCTTGAAATGAAAACGCGTCAGGCCT 6654

Qy          5407 ACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTC 5466
                 ||  |||  |||| ||| | || || || |||   ||||||||||| || |||||    |
Db          6655 ACGCCGGTTCGCAAATGCACGAGCCCGCTTTTGCCGCGTTTCAGGAGGTCGCGATGGCGC 6714

Qy          5467 ACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCA 5526
                 |  |  |  | |        ||||| || ||||||||||| || || |||||||  |||
Db          6715 ATGATATCGCTCCCGCCTACGCGTTCGACCATCTGGAAGGTTTTGCCATGGATGTGCGCG 6774
```

```
Qy       5527 ACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGG 5586
              |    |  |||  ||   | || || ||||||||||| ||||||||||||||||| || || |
Db       6775 AAACGCGCTACCTGACACTGGACGATACGCTGCGTTATTGCTATCACGTCGCCGGTGTTG 6834

Qy       5587 TCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCT 5646
              | ||  ||||||||||||   | |||||||| |||||  | ||   ||||||||| |||||
Db       6835 TGGGCCTGATGATGGCGCAAATTATGGGCGTTCGCGATAACGCCACGCTCGATCGCGCCT 6894

Qy       5647 GCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCG 5706
              ||||| | || ||||| |||||| | || |||||||||||| || ||||| || |||||
Db       6895 GCGATCTCGGGCTGGCTTTCCAGTTGACCAACATTGCGCGTGATATTGTCGACGATGCTC 6954

Qy       5707 AAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATA 5766
              |    || ||||| ||||||||  ||  ||||||| ||  |||  ||| |       |  |
Db       6955 AGGTGGGCCGCTGTTATCTGCCTGAAAGCTGGCTGGAAGAGGAAGGACTGACGAAAGCGA 7014

Qy       5767 CGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGG 5826
                    |||| ||  || | ||    ||  |   |    | || | |||  | ||   ||
Db       7015 ATTATGCTGCGCCAGAAAACCGGCAGGCCTTAAGCCGTATCGCCGGGCGACTGGTACGGG 7074

Qy       5827 AGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGG 5886
              | ||||||||||||||    ||  |     ||||| |  |    |||||  | |||||||
Db       7075 AAGCGGAACCCTATTACGTATCATCAATGGCCGGTCTGGCACAATTACCCTTACGCTCGG 7134

Qy       5887 CGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACG 5946
              | ||||||||||| || ||       || || ||  ||||||||||| |||||| | || |
Db       7135 CCTGGGCCATCGCGACAGCGAAGCAGGTGTACCGTAAAATTGGCGTGAAAGTTGAACAGG 7194

Qy       5947 CCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGC 6006
              |||||  ||| |||||||||   || |||  |||    |  |  |||||| |  ||||
Db       7195 CCGGTAAGCAGGCCTGGGATCATCGCCAGTCCACGTCCACCGCCGAAAAATTAACGCTTT 7254

Qy       6007 TGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGG 6066
              || |||  |   | |||  |||  | ||||| ||  ||        |  ||| ||||| |
Db       7255 TGCTGACGGCATCCGGTCAGGCAGTTACTTCCCGGATGAAGACGTATCCACCCCGTCCTG 7314

Qy       6067 CTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGGCTTG 6126
              ||  ||| ||||||||| ||        | ||    | || |  |  | |||||| || ||
Db       7315 CTCATCTCTGGCAGCGCCC------GATCTAGCCGCATGCCTTTCTCTCAGCGTCGCCTG 7368

Qy       6127 CAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCCCGCG 6186
               || |||   | |||||||||||| || |||||||| |||||||||||| ||    || |
Db       7369 AAGTTTAGATAACGGTGGCGCGTACAGAAAACCAAAGGACACGCAGCCCTCTTTTCCCCT 7428

Qy       6187 CACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCGGGAT 6246
               || ||||||||||||| |||||| |||||||||||| || || || |||||||||||| ||
Db       7429 TACAGCATGATGCATACGGTGGGCCATGTATAACCGTTTCAGGTAGCCTTTGCGCGGTAT 7488

Qy       6247 ATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCGCGCC 6306
               ||||||||||||||||| || || ||||| || ||||| ||||| || || ||    ||
Db       7489 GTAGCGGAACGGCCAGCGCTGGTGTACCAGTCCGTCGTGGACCATAAAATACAGTAAACC 7548

Qy       6307 GTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATA 6366
               || |  ||||| || |||||||||||||| |||||||| || |||      ||||  ||
Db       7549 ATAAGCGGTCATGCCTGCACCAATCCACTGGAGCGGCCAGATTCCTGTACTGCCGAAGTA 7608

Qy       6367 AATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTT 6426
              ||||||  ||||||| ||  |  || || |||| ||||| |||||||| |  |||||
Db       7609 AATCAGGGCAATCGACACAATGGCGAATACCACGGCATAGAGATCGTTAACTTCAAATGC 7668

Qy       6427 ACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTA 6486
               ||  |  |||||||||| || || ||||||||||||||| ||||| ||||||||||||||
Db       7669 GCCTTTACGCGGTTCATGATGTGAAAGATGCCAGCCCCAACCCCAGCCGTGCATGATGTA 7728

Qy       6487 TTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCAC 6546
              |||||| | ||| || || || | ||||||  | ||||||||  | |||| ||| ||  |
Db       7729 TTTATGTGCCAGTGCAGCAACCACTTCCATGCCGACCACGGTGACAAACACGATCAGGGC 7788

Qy       6547 GTTCCATAACCAGAGCATTGTTCGTCCA 6574
               ||||| | ||| | |||  ||  || |
Db       7789 ATTCCAAATCCACAACATAATTTCTCAA 7816
```

```
RESULT 7
US-10-808-807-18
; Sequence 18, Application US/10808807
; Publication No. US20040253663A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. du Pont deNemours and Co., Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  APPLICANT:  Sedkova, Natalia
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2365 US NA
;  CURRENT APPLICATION NUMBER: US/10/808,807
;  CURRENT FILING DATE:  2004-03-24
;  PRIOR APPLICATION NUMBER: US 60/477,874
;  PRIOR FILING DATE: 2003-06-12
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
;   LENGTH: 8814
;   TYPE: DNA
;   ORGANISM: Pantoea agglomerans strain DC404
US-10-808-807-18

  Query Match             30.6%;  Score 2138.6;  DB 18;  Length 8814;
  Best Local Similarity   60.7%;  Pred. No. 0;
  Matches 3811;  Conservative    0;  Mismatches 2279;  Indels  185;  Gaps   12;

Qy        309 ACCATGGATAGCCATTATGACCACCCATGTCGACACCACAGCACATCAGACAAGCGAACT 368
              ||   || || | |  |||| |  |   | || ||     |   |||     | |||||
Db       2384 ACGCCGGGTACCAACCATGACAAGACCCTTTGAAACACATCCCGGTCACGACGGGGAACT 2443

Qy        369 CCTTCAGCTGCAGCAAATTTTACAGGCGCATCTTGAACATTTACTGCCTGCCGGACAGCA 428
               | | |||||||        | ||    |  || ||  |  | ||||| |  ||  |  |
Db       2444 GCATGAGCTGCACGCTGCCCTGCAACGTCGCCTGGATGAACTGCTGCCCGTTGGCGATGA 2503

Qy        429 AAGCGATCGCGTGCGTGCCGCGATGCGTGCCGGAACGCTGGCGCAGGGCAAACGTATTCG 488
                | ||||| ||  |   ||| ||||| |  ||    |||||| | ||| ||||| |||||
Db       2504 GCGGGATCGGGTCAGCAGCGCAATGCGCGAAGGCGTACTGGCACCGGGGAAACGCATTCG 2563

Qy        489 TCCTTTATTACTGCTGCTGGCAGCGCGCGATATGGGTTGCGAGCTGACGCAAAATGGCGT 548
               ||  |  | ||| | || || || |||||  | || |||| |     ||    ||| |
Db       2564 CCCGCTGCTCCTGATCCTCGCCGCCCGCGACCTCGGCTGCGATCGCGACCACCCCGGCCT 2623

Qy        549 TCTCGATCTCGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACAT 608
               || ||| | |||||||| |||||||||||||||||||||    |||||||| || || || ||
Db       2624 GCTGGATATGGCCTGTGCGGTGGAAATGGTGCACGCCTCGTCGCTGATCCTCGACGATAT 2683

Qy        609 TCCCTCGATGGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGG 668
              |||||  |||||||||||||  | | || || ||||||||||| | ||||| | || | |||
Db       2684 TCCCTGCATGGATAACGCGGCGCTCCGGCGCGGTCGCCCTACCATTCATCGCCAGTATGG 2743

Qy        669 TGAAAACGTGGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGC 728
              |||| ||||||| ||||||||| ||  | ||| |||| ||| |||| ||||  ||||| |
Db       2744 TGAAGACGTGGCAATTCTCGCTGCGGTAGCGTTGCTCAGCAGCGCCTTTGGCGTGATGGT 2803

Qy        729 CATTGCACCCGGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGC 788
              |   || |  || ||| || ||             | ||| | || || || |    ||
Db       2804 CGCGGCGCAGGGATTGTCTCCCGAGTGCCGCAGCCAGGCGGTGGCGGAGCTGTCGATGGC 2863

Qy        789 CGTCGGCCTGCAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAG 848
               |||||    |||||  | ||||| || ||  |   ||||||||||  || |||||
Db       2864 GGTCGGTACCCAGGGTCTGGTGCAGGGTCAGTATAAGGATCTGCGTGAAGGCACCGCCCC 2923

Qy        849 CCGCAGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGC 908
               |||||| | || | |||||||||  |||||||||||||||||||||||||||||| | ||
Db       2924 GCGCAGCGCCGAGGAGATCGCCACCACCAACGAACTGAAAACCAGCGTGCTGTTTGGTGC 2983

Qy        909 CACGCTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAG 968
              ||||||||||||| |||||  | || || | ||| || |||  || ||| || | | |  |
Db       2984 CACGCTGCAAATCGCGGCCCTGGCGGCAGGCGCCTCGCCGGCGGCGCGCCAGAAAATGCG 3043
```

```
Qy     969 CTTCTTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTG 1028
           || ||| || |||||||| ||||||||||| || ||||| ||||| || || |||||
Db    3044 CTGCTTTGCGCAGGATTTAGGCCAGGCGTTCCAGCTGCTGGACGATCTGGCGGACGGCCA 3103

Qy    1029 CAAACACACCGGTAAAGATGTGCACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCT 1088
                  ||||| |||||  |  |  ||||   ||| || |||||||||||   ||||||
Db    3104 TGCCGGGACCGGCAAAGACATCAATAAGGACGCGGGTAAGTCCACGCTGGTGGCGATGCT 3163

Qy    1089 CGGTGCTGACGGCGCGGAACGTCGCCTGCGCGATCACCTGCGCAGCGCAGATGCACACCT 1148
           |||    ||||  | |      || ||   |   || |||||| ||||||| || ||  |
Db    3164 CGGCAGCGACGCGGTGCGCGAGCGGCTCGACACCCATCTGCGCCGCGCAGACGCCCATTT 3223

Qy    1149 TGCCTGCGCCTGCCATCGCGGCATCGCCACTCGCCAATATATGCACGCGCTGTTTAATCA 1208
           | |  ||||||||        |   ||||| || |  | |||||||||   ||||    |
Db    3224 TTCACGCGCCTGCGGAAAAAACCAGGCCACGCGACGCTTTATGCACGCCTGGTTTTCAAA 3283

Qy    1209 ACAGCTAGCGATATTCAACTGAGCGCGGCTCAGCCGGTGGGCCACTTTGCGGTGATCGCG 1268
           |||||| ||    || | ||||||   |   |  |   ||     |||| || |||| |
Db    3284 ACAGCTGGCCGCGTTTAGCTGAGCAACGGATACACCCCGGTAATATTTGTGGAGATCACA 3343

Qy    1269 CCGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCGTTAGCACAAACGCTGCTGGCGCGC 1328
                                                                     |
Db    3344 T---------------------------------------------------------GA 3346

Qy    1329 GGCCATCGCATCACATTCATCCAGCAAGCCGATGCCCGCACTTTGCTTAGCGACGAACGC 1388
            |    ||||||   |||| |    ||       || | |  | |    ||  |     |
Db    3347 AGGACGCGCATCTGGTTCAGCGTAAAAATGACCACCTGGATATCGTGCTGCACCCTGACC 3406

Qy    1389 ATCGATTTTGTTGCCGTCGGCCAACAGACGCATCCTGCCGGTTCGCTGGCGCCCGTGTTG 1448
                 |  ||  |  |||  |   |   |   |||| || || |    |  |
Db    3407 GGGCGATGAGTACCATTCGCACCGGATTTGACGCCTGGCGTTTTGAACACTGC------- 3459

Qy    1449 CATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTTCGCGTGATCGACGATCTCGCGTCC 1508
                             | |  || |  ||||  || ||       ||||||| |  ||
Db    3460 ------------------GCCCTCCCGGAGCTGGATCTCGACGGTATCGATCTCTCCACC 3501

Qy    1509 TGCACCGATATGCTGTGCCGCGAACTGCCTGCGGTACTGAAAGCATTGAACATCGATGGC 1568
             |     |   |     |   ||   || | | |  | |   |  ||| |  ||  |
Db    3502 ACCCTGTTTTCCCGCCCGCTGAAAGCCCCGGTGCTGATCAGCTCCATGACCGGCGGCGCG 3561

Qy    1569 GTGATCGCCGACGAAATGGAAGCGGCGGGCGGATTGGTCGCTGAAGCGCTGCATCTGCCG 1628
           | |  ||||   || ||  |           |    ||  ||  || | ||    ||| ||
Db    3562 GCGCGCGCCAGAGACATTAACCGTCATCTGGCCCAGGCGGCGCAAACCCTTGGGCTGGCG 3621

Qy    1629 TTTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGTGAAGCCGGGATTCCGCTTGCGGTG 1688
            | |     |   ||     | || || |  |    |  ||   |     ||| |  |
Db    3622 ATGGGCGTCGGTTCCCAGCGCGTGGCGCTGGAGGACGGCGCGCAGCACGGGCTGGATGCC 3681

Qy    1689 ATGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTGAAACGTTTTCAGGCCAGCAGCGAT 1748
             ||    || | | || | ||| |       ||||                     | |
Db    3682 CAGCTACGCCATATCGCCCCGGACGTGCCGCTGCTG---------------------GCT 3720

Qy    1749 ATCTATGATCGCATCATGCGTCGTCACGGCGACGTGATCCTCAAACACGCGCGGGCGTTT 1808
           | |  ||          |   ||    |   | | | |   | |   |  ||| ||| |
Db    3721 AACCTTGGCGCGGCGCAGATCCGCGGTGCGCAGGGGCTGGACTACGCCCGGCGCGCGGTG 3780

Qy    1809 AATTTGACGGAGCGGCGCGGATTACATCAGTGCCTGTCGCCGCTGGCACAAATCAGCCAG 1868
            |  |||  ||      ||  |||  |  |   |||   ||||||            |||
Db    3781 GACATGATCGACGCCGACGCGTTAATTGTGCATCTGAACCCGCTG------------CAG 3828

Qy    1869 ATGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTGCCCGCCTGCTATCACGCCGTGGGG 1928
             || ||        |        ||  |    |    |  |  | |   |   | |  |
Db    3829 GAGGCGCTCCAGGGCGGCGGCGATCGCGACTGGCGCGGCATCCTCAACGCCATTGCGCAG 3888

Qy    1929 CCACTCCGCGCCCCGGTTTCTCCTGCGCCGCTCCATGCGCCCTGGCCAGCGCTGCGTCAG 1988
           |   | |||| || |      || | |  |  |  | | |   ||| |  ||     ||
Db    3889 CTGGTGCGCGACCTGCCGGTACCGGTGGTGGTTAAAGAG--GTGGGCGCCGGGATCTCCC 3946

Qy    1989 CCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGCCATCGCTTCCGGCTGTTTCTGCAT 2048
           | |  | |   ||||     | | |||  |    | |   | |      ||
Db    3947 CGGACGTTGCCTGCCGACTGGCGGACGTCGGCGTGGCGATGATCGACATTG--------- 3997

Qy    2049 CTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTGGTGATCGCCCATTGTGGGGGATTA 2108
```

```
           | ||||| ||||   |       |||||  | |  | ||| ||| |   |
Db    3998 CCGGCGCGGGCGGAACCAGCTGGGCGGCGGTGGAAGCTGAACGCGCCCCG---------- 4047

Qy    2109 AACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGCGCGGCGTGGGTGACGGATTTCGTC 2168
                ||   || |||   |  ||| | | ||    || |  ||||  |    |   | |
Db    4048 ---ACCCCCGAGGCGCGAAATGTGGCGATGGCCTTTGCCGACTGGGGCATTCCTACTGCC 4104

Qy    2169 GATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTTATCACTCATGCCGGGTTAAACAGC 2228
           |||  ||       | | || |  |||| ||    |||| | | |  ||  |     | ||
Db    4105 GATGCGCTGCGTCGCGTCCATCTTGCGCTGCCTGATATCCCGCTTATCGCCTCCGGCGGC 4164

Qy    2229 GCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTGGCGCTGCCGATTGCTTTTGATCAG 2288
                  | |   | ||    |  |           ||  |  || | | |   |||| |
Db    4165 ATCGCCAACGGCATTGACGCAGCAA---------AAGCCATCGCGCTGGGTGCAGATCTG 4215

Qy    2289 CCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTTGGTCGCCGCGCATCACGCTTTAGC 2348
              |||  ||| |||    |   |  ||||| |   |    | |||         ||  |
Db    4216 GTGGGCCAGGCCGCGGCGGTGCTGGCGCATGCCAACGCCTCCGGCGACGCGGCAATTGCC 4275

Qy    2349 CGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTGCTGACCGACGATCGTTACGCGCTA 2408
           | | | |     |||     ||| ||||       |     |     ||  |||  ||
Db    4276 CATTTCCGCACCCTGATTACGCAGCTGC-------GGATCGCCTGTTTCTGTACCGGCAG 4328

Qy    2409 CGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCAGGCGGTTGCCAGCGTGCCGCCGAC 2468
            | |   | ||       |||   |  ||    ||   ||||   | |||   |  |
Db    4329 TGCAAACCTGCAGGCGTTGCGACACGCCACGCTGCTTCCGGTCAACGGCGGCGCATCCCT 4388

Qy    2469 ATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCA 2528
            |     || | |     |   | | |  ||  ||                        |
Db    4389 GTGACGCAGTACGGTGCCTTATACCGGGGAGCGGT---------------------ATGA 4427

Qy    2529 CGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTC 2588
              |||  ||| ||||| ||||||| || || ||||| || ||  | || ||   ||| |
Db    4428 AAAAATGGGATCTGATTCTGGTCGGCGCGGGGCTGGCCAACGGGCTTATCGCCTGGCGAC 4487

Qy    2589 TGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAG 2648
           |    ||    || ||    ||        |  ||||||||| |||  |     || |  |
Db    4488 TAAAGCAGCGTCATCCGACGCTTGCTGTATTAATGCTGGAGTGCGGCGACGCGCCCGGCG 4547

Qy    2649 GCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGC 2708
           | || || ||||||||| ||||| ||   |||| |||   | |  || |    |||||||
Db    4548 GAAACCACACCTGGTCCTTTCACCAACACGATATCACGCCAGCCCAGCACGCCTGGCTGG 4607

Qy    2709 AACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCA 2768
             |||||| |  ||   || |||   || ||  | || |  |||||     ||   ||||
Db    4608 CGCCGCTGGTGGCCCATCGCTGGGACGGGTACGACGTCCACTTTCCGAACGTGTCGCGCA 4667

Qy    2769 ATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGG 2828
             ||| | |  |  ||     ||||| | ||      ||||||||      |   || |
Db    4668 CCCTGCATGACGGCTACCTGACCATCACCTCCACGCGTTTTGCCCAAGCGATGCGCGGGC 4727

Qy    2829 CGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGA 2888
            ||||   ||  || ||  |||||||    ||||   || ||| |       | | || |
Db    4728 TGATGAAAGAGAATTTGCTGACAAACGTGACCGTGTCACGGGTGAGCGGGCAGGAAGTAA 4787

Qy    2889 CGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGC 2948
           | ||    || || ||     || | ||    | ||||||||||||| ||||||   ||||
Db    4788 CCCTCAGCGACGGACGACGCTTTACCGCCGGGGCGGTGATTGATGGCCGCGGCTATCAGC 4847

Qy    2949 CGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGC 3008
           |  |||| || ||     | || ||||||| |||  | || || |||||||| ||| |
Db    4848 CCTCGCCGCACCTCAGCATTGGCTATCAGGCGTTCATCGGCCAGGAGTGGCAACTGACCG 4907

Qy    3009 AGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTT 3068
            ||| |||||  |   ||  ||||||||||||||||||  |||||  |||      || |
Db    4908 CGCCCCACGGGTTAACGCGCCCGATCCTGATGGATGCCCGCGTCGCCCAGGGCAACGGCT 4967

Qy    3069 ATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATT 3128
           | || |||||||| || |||||||||||||||    ||  | || ||||| || || |
Db    4968 ACCGCTTTGTCTATACCCTGCCGCTCAGCGCCGACACCCTGCTTATCGAAGACACGCACT 5027

Qy    3129 ACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCA 3188
           || || ||   ||  |||| |  |   |  | || ||   ||  || ||||| || |||
Db    5028 ACATTGACGGCCCGACGCTCGACGCCGATTCAGCCCGCGCGCGGATTGCCGATTACGCCC 5087

Qy    3189 ATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTA 3248
             ||||| ||||||  |||      |||| ||||||| || || ||     | ||||| |
Db    5088 GCCAGCAGGGCTGGCAGCTTGCGCGGCTGGTGCGTGAGGAACAGGGGGCGCTGCCGATCA 5147

Qy    3249 CCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCC 3308
           |||||  |||| |    |     |||||||| |||   |     ||  ||  ||||||||
Db    5148 CCCTGTCCGGCGATCCGGCCGCCTTCTGGCACCAGTTCCATCATCAGCCGGTCAGCGGCC 5207

Qy    3309 TGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAG 3368
           |||||||||| ||||| |||||||||||||| || ||  |||||   || ||   ||| |
Db    5208 TGCGCGCCGGTCTGTTCCATGCCACCACCGGCTATTCGCTGCCGCTGGCGGTTCGGCTGG 5267

Qy    3369 CGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAAC 3428
           ||||     | ||      |  |       ||     | |||||   || |  ||||
Db    5268 CGGACCGCATTGCCAACGCGCCGGGACTGCATCAGGGCGCGCTCTATCAGCTGATCGCCG 5327

Qy    3429 GCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGT 3488
             || ||    |  || ||||    ||| || |||||||| ||||| ||||||||||| |
Db    5328 ATTTCGCGGCGCGCCACTGGCAGACACAACGCTTTTTCCGCCTGCTTAACCGCATGCTTT 5387

Qy    3489 TTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATG 3548
           |  ||||||| |  ||  | |||||||||||||||||||| || ||||||| ||| || |
Db    5388 TCCTGGCCGGCACACCCGACCAGCGCTGGCGCGTGATGCAGCGGTTTTACCAGCTTGACG 5447

Qy    3549 CCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTC 3608
               | | ||   ||| ||||| ||||| || || |||     |||     ||||  | |
Db    5448 AGCAGCTGATCGCCCGTTTTTATGCCGGCCAGCTTCGCTCCGCCGACCGCGCGCGCCTGC 5507

Qy    3609 TGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTT------GAATT 3662
           ||   ||||| ||||||||||| || |  |  ||| |    || ||| |        |||
Db    5508 TGCTTGGCAAACCGCCGGTGCCGATTGTCGGGGCGATCAAAGCCCTGCTCCACACTCATT 5567

Qy    3663 CTGTCGAACCAGGGAAGAAAAAATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGG 3722
           ||      | ||   |  | |||||||||  ||   ||| |||||||| || || |||||
Db    5568 CTTCTCTGCGAGCCCATCATAAATGAAACAAACCATTGTAATTGGCGCCGGGTTCGGCGG 5627

Qy    3723 CCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCG 3782
            |||||||||||||||||||| || ||||||||||| || |||||  | || |||   ||
Db    5628 ACTGGCGCTGGCGATTCGCCTCCAGGCGGCGGGCATTCCTACCACGCTGCTGGAGAGCCG 5687

Qy    3783 CGACAAACCGGGCGGACGCGCCTATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGG 3842
           ||||||||| ||||| ||||||||||| |  || ||  | ||||||||||| ||||| ||
Db    5688 CGACAAACCCGGCGGCCGCGCCTATGTCTACGAAGATCGCGGCTTTACCTTTGATGCGGG 5747

Qy    3843 ACCCACGGTGATCACCGATCCCAGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATC 3902
            ||||| || ||||||||||||  |||||| || ||| ||||||| || || ||||||
Db    5748 TCCCACCGTCATCACCGATCCCTCCGCCATTGAGGAGCTGTTCACCCTCGCCGGAAAACG 5807

Qy    3903 GCTCAGCGATTACGTCGAGCTGATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGA 3962
           ||| |  || ||||| |||||||||||||| ||||| |||||||||||||||||||||||
Db    5808 GCTGAAGGACTACGTTGAGCTGATGCCGGTGACGCCGTTCTATCGCCTGTGCTGGGAAGA 5867

Qy    3963 TGGCAAACAGCTTGATTACGACAATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTT 4022
            |||||     | || |||| |||  ||||| ||  ||| |||  |||||||||| ||||
Db    5868 CGGCAAGGTTTTCGACTACGCCAACGATCAGGCGGCGCTTGAGTCGCAGATCGCCGCGTT 5927

Qy    4023 CAATCCGCAAGATGTAGAAGGCTATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAG 4082
            || ||| | || || |  |||||||  |  || || | ||| || || |  || |||
Db    5928 TAACCCGAACGACGTGGCGGGCTATCACCGCTTCCTCGACTACTCCCGGGCGGTGTTTGC 5987

Qy    4083 AGAGGGTTATCTGAAACTCGGCACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGT 4142
            || || |||||||| |||||| |||||||||||||   | | || |||||||||||||
Db    5988 CGAAGGCTATCTGAAGCTCGGCGCGGTGCCGTTTCTCTCGTTTCGCGACATGCTGCGCGC 6047

Qy    4143 CGCGCCGCAGTTGGGACGTCTGCAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATT 4202
           ||  || ||  |||  || ||||| |||||||||||||| |||  ||  ||| ||   |
Db    6048 CGGTCCTCAACTGGCGCGGCTGCAGGCATGGCGCAGCGTGTACGACAAAGTGTCGGCCTA 6107

Qy    4203 TATTCAGGACGATCATCTGCGTCAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAA 4262
             |  | ||||| || ||||| ||||| ||||| || |||||  ||||||||||||| ||
Db    6108 CGTGGAAGACGAGCACCTGCGGCAGGCATTTTCGTTTCACTCGCTGCTGGTGGGCGGCAA 6167

Qy    4263 TCCTTTTGCAACGTCATCGATCTATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGT 4322
            || ||  | ||||| || || || ||| | || || || |||||||| |||||||||||
Db    6168 CCCGTTCTCCACGTCTTCTATTTACACCCTGATCCACGCCCTGGAGCGGGAATGGGGCGT 6227

Qy    4323 GTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGA 4382
            ||||| ||||||||||||||||| |||||||| |||||||||| |   |||||  ||||
Db    6228 CTGGTTCCCGCGCGGCGGCACCGGTGCGCTGGTTCAGGGCATGGTGAAGCTGTTTCAGGA 6287

Qy    4383 CTTGGGCGGCGAGCTGTTACTGAATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCG 4442
             | ||||||   ||    || || ||  | ||    | |||||| ||    | |||||
Db    6288 TCTTGGCGGCACCCTCACCCTTAACGCTCAGGTTGAGCGGCTGGAGACGGTGGACAATCA 6347

Qy    4443 CATTAGCGGCGTTCAGTTAGAGGGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAA 4502
             | |  | ||| ||  | |    ||| |  ||  | || || || || |||||||| ||
Db    6348 GGTGAAGGCCGTGCATCTGGTTAACGGGCAGCGGCTGGAGGCTGCGGCGGTGGCCTCGAA 6407

Qy    4503 TGCCGACGTGGTGCATACCTACGACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACG 4562
            || ||||||||  ||||||| | |  ||||||  |||| || ||||  |
Db    6408 CGCGGACGTGGTAAATACCTATGCCCGACTGCTCGGCCATCACCCGCACGGCGCCGCTAC 6467

Qy    4563 TGCGACATCGCTGAAGCGTAAGCGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCT 4622
            || | |  |||||| || ||||||||||||||||||||||| || ||||||||||||||
Db    6468 GGCCAAAAAGCTGAAACGCAAGCGCATGAGCAACTCGCTGTTCGTGCTCTATTTTGGCCT 6527

Qy    4623 GAATCAGCCGCATGAACAGCTCGCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGA 4682
           | |||| |  ||    ||||| |||||||| |||||||| |||||||||||||||   |
Db    6528 GGATCACCATCACACCCAGCTGGCGCACCATACCGTCTGCTTTGGCCCGCGTTATAAAGC 6587

Qy    4683 GTTGATCGATGAGATTTTCAACAGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCA 4742
           | | |||||||| ||||||| |  |  |   ||| | || |||||||| || || |||||
Db    6588 GCTAATCGATGAAATTTTCAGCGCCGACACCCTGTCGGAAGATTTTTCGCTCTATCTGCA 6647

Qy    4743 CGCGCCCTGCAGCAGCGATCCGTCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTT 4802
            |||||||||   | ||| ||||||||||| ||||| || ||||||||||  |||||| |
Db    6648 TGCGCCCTGCGTAACCGACCCGTCGCTGGCCCCGCCGGGGTGCGGCAGCTACTATGTGCT 6707

Qy    4803 AGCGCCGGTGCCGCATCTCGGCACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTT 4862
            |||||||||||||| ||||| | |||     ||||||||     |||||| |||||  |
Db    6708 CGCGCCGGTGCCGCACCTCGGTAACGCCCCGCTCGACTGGAGCGTGGAAGGGCCGCGTCT 6767

Qy    4863 GCGCGATCGAATTTTTGCTTATCTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATT 4922
           ||| ||||| ||||||| |||||| ||   || ||| ||||||||  | ||    ||  |
Db    6768 GCGGGATCGCATTTTTGATTATCTCGAAGCGCGCTATATGCCGGGGCTGCGCTCCCAGCT 6827

Qy    4923 AGTGACACACAGAATGTTTACGCCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGG 4982
            ||||| ||| | ||||| ||||||   ||||||||||| |||||  |||||   || ||
Db    6828 GGTGACGCACCGCATGTTCACGCCGGAAGATTTTCGCGATACGCTCGATGCCTGGCAGGG 6887

Qy    4983 CTCGGCGTTTTCGCTGGAGCCGATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCG 5042
            || |||||||| |||||||||||  | || || |||||||||||||| ||||| |||||
Db    6888 GTCAGCGTTTTCACTGGAGCCGATCCTCACCCAGAGCGCCTGGTTCCGGCCGCACAACCG 6947

Qy    5043 CGATGCCGATATCAGCAATCTCTATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGT 5102
           |||   ||   |    || ||||| ||||| || ||||| |||||||| ||||| |||||
Db    6948 CGACAGCGTGGTTGATAACCTCTACCTGGTCGGCGCCGGAACGCATCCCGGCGCTGGCGT 7007

Qy    5103 GCCCGGCGTGATCGGTTCGGCCAAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGA 5162
           ||| ||||||||||| || |||||||| || |||  | | ||| |  |||||   ||  |
Db    7008 GCCGGGCGTGATCGGATCCGCCAAGGCAACGGCCCAGTTAATGTTAAAGGATTTAGCGTA 7067

Qy    5163 ATGAATCGACAGCCTTTACTTGAGCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGT 5222
           |||      |||||  | || || || |  |     |||||| | |  || || || |||
Db    7068 ATG---TCCCAGCCGCTTCTCGAACACGCCAGCGCCACCATGACCGCCGGTTCTAAAAGT 7124

Qy    5223 TTCGCCACCGCCGCCAAGCTGTTTGATGCACCGACGCGCCGCAGCACGCTGATGCTGTAT 5282
           |||||||||||| | |||||||||||   ||  || || |||||| |||||||||| |||
Db    7125 TTCGCCACCGCCTCAAAGCTGTTTGACAAACGCACCCGGCGCAGCGCGCTGATGCTCTAT 7184

Qy    5283 GCGTGGTGTCGTCACTGCGATGATGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACG 5342
            | ||||| ||  ||||||| ||||| || || || ||   | ||||    |  | | ||
Db    7185 ACCTGGTGCCGCTACTGCGACGATGTTATCGACGGACAGGTGGTGGGTTTTGCTGCCCCG 7244
```

```
Qy          5343 CAGCATGCCGTCGAAGACGCGCAGGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGC 5402
                     |      |||       |||| ||  |||| |  ||||  |     || ||||||
Db          7245 ACCGAGCAGAGCGACACGCCCGAGGCGCGCCTGCAACGGCTGCGTAAGATGACGCGCCGC 7304

Qy          5403 GCCTACAGCGGCGCGCACATGGATGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATC 5462
                 ||||||  ||| |    ||||| | || ||| ||||    || ||||||||| || ||  ||
Db          7305 GCCTACGACGGGGAAACCATGCAAGAGCCGCCGTTCGCCGCCTTTCAGGAGGTTGCCCTC 7364

Qy          5463 ATTCACCAGCTGCCGCAACAACTGGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCA 5522
                    ||     | ||||     | ||| || || || ||||||||||  || ||||| |
Db          7365 GCCCATGCCATTCCGCCTACTCAGGCCTTCGACCACCTGGAAGGCTATGCGATGGACGTG 7424

Qy          5523 CGCAACGAACATTACGCGAGCTTCGATGACACGCTGCGTTACTGCTATCACGTCGCGGGC 5582
                 |||||||| |  ||    ||| ||||||| ||||| || ||||| |||| ||||||| ||||||
Db          7425 CGCAACGAGCGCTATTACAGCCTCGATGATACGCTCCGCTACTGTTATCACGTGGCGGGC 7484

Qy          5583 GTGGTCGGTTTGATGATGGCGCGCGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCAC 5642
                 ||||||||  ||||||||||  | || ||||| ||||| ||||||||   ||| |||| |
Db          7485 GTGGTCGGCCTGATGATGGCCAGGGTGATGGGAGTGCGGGACGAAGCCACGCTGGATCGC 7544

Qy          5643 GCCTGCGATTTAGGACTGGCGTTCCAGCTCACTAACATTGCGCGCGACATTGTAGAAGAT 5702
                 ||||||||| | ||  | || || ||||||||| || || || ||  | || || || || |||
Db          7545 GCCTGCGATCTGGGCATTGCCTTTCAGCTCACCAATATCGCCAGGGATATCGTTGACGAT 7604

Qy          5703 GCCGAAAATGGTCGCTGCTATCTGCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCC 5762
                 ||  |    || ||||||||| ||||||||    ||||| |   | |  ||| |    |
Db          7605 GCGCAGGTGGGACGCTGCTACCTGCCGCAGCAGTGGCTGGCGGAAGTCGGACTCAATGAA 7664

Qy          5763 GATACGCTGACTGCACCGCAACATCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTG 5822
                  | ||    || |  |  |   | ||| | ||||| ||    |||||||||||  | |||
Db          7665 CAGACCTGCACCGTGCGGGCCAACCGTCCGGCGCTGGCGCGTCTGGCAGCGCGGCTGGTG 7724

Qy          5823 GCGGAGGCGGAACCCTATTATCACTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGC 5882
                  | ||||| || |||||||||||| || ||||   ||||  |   || | | || ||||||
Db          7725 ACCGAGGCTGAGCCCTATTATCAGTCAGCGCTTGCCGGGCTGGGGGATCTGCCCCTGCGC 7784

Qy          5883 TCGGCGTGGGCCATCGCTACGGCTCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAG 5942
                 || || ||||| || || || || | ||| || ||||| || || || || || || | |
Db          7785 TCCGCCTGGGCGATTGCCACCGCGCACGGGGTGTATCGTGAGATCGGGGTGAAGGTGCTG 7844

Qy          5943 CACGCCGGTGTGCACGCCTGGGATTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCG 6002
                    || ||||   | || ||||| | || ||| |||| |     |  || || ||||||
Db          7845 ATGGCGGGTGAAAAAGCATGGGATACCCGCCAGGGCACGACGCGCGCGGAGAAGCTGGCG 7904

Qy          6003 CTGCTGGTGAAAGGGGCAGGTTTGGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGT 6062
                 ||| |  |    || ||      ||||||  |||| ||   | |  |        |||||
Db          7905 CTGGTTATTTCCGGCGCGAAGCAGGCGATGGCTTCCCGGAAGGCGAGCTGGCCGCCGCGC 7964

Qy          6063 CCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTTTACGTCCGTGACGCTGGCGCAGCGTGG 6122
                     |    || |||||||| || || |       |  |||| ||| | ||||||   |
Db          7965 GATCCGCACCTCTGGCAGCGCCCGCGCTAGCGGGTCTGCCGTTACGTTCGCGCAGCACCG 8024

Qy          6123 CTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGAAACCAAACGACACGCAGCCTTCACGCC 6182
                 | |||||||| | || |||||| |||||||  || || ||  || ||||| ||||| ||||
Db          8025 CCTGCAGCTTGTCCACCGGTGGGGCGTAGATAAACCCGAAGGAGACGCACCCTTCGCGCC 8084

Qy          6183 CGCGCACCGCATGATGCATGCGGTGCGCCATGTATAAGCGCTTAAGATAGCCTTTGCGCG 6242
                 | ||||||||  |||||||  ||||| |||||||| | |||    ||||||||   |||||
Db          8085 CCCGCACCGCGTGATGCAGCCGGTGTGCCATGTAGAGGCGGCGCAGATAGCCGCGGCGCG 8144

Qy          6243 GGATATAGCGGAACGGCCAGCGTTGATGCACCAGGCCATCGTGCACCATGAAGTAGAGCG 6302
                 | |  || ||||||||||||||| || || || |  ||||||||| || || ||||||| |
Db          8145 GCACGTAACGGAACGGCCAGCGCTGGTGGACTAAACCATCGTGAACGATAAAGTAGATCA 8204

Qy          6303 CGCCGTACGTCGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGA 6362
                 |||||||    |||||||| || ||||||||||||| |||||||||    ||||     ||
Db          8205 CGCCGTAGCCGGTCATTCCCGCGCCAATCCACTGAAGCGGCCAGTACCCTTCGCTGCCCG 8264

Qy          6363 CATAAATCAGCACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAA 6422
                 | |||||||| | |||| |||||| ||||||||||||||||||||| ||||||||  ||||||
Db          8265 CGTAAATCAGCGCAATGGCCAGTAGCGCAAACACCACCGCATAGAGATCGTTACGCTCAA 8324

Qy          6423 ACTTACCGCTGTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGA 6482
                 ||   ||  || ||||   |||||||||  ||||||||||||||||||| ||||||||||
Db          8325 ACGCCCCTTTGCGCGGGGTATGGTGCGAATGATGCCAGCCCCATCCCCAGCCGTGCATGA 8384

Qy          6483 TGTATTTATGCGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAA 6542
                 |||| || || |  | ||  || ||   |||||| |  |  |  ||    | || |||
Db          8385 TGTACTTGTGTGCGAACGTTGCCACCCCTTCCATGATGATGATAGTCAGTAGCACGATCC 8444

Qy          6543 GCACGTTCCATAACCAGAGCATTGTTCGTCCATTT 6577
                      ||||| |||   |||||| | |  | | | |
Db          8445 CGGTATTCCACAACGCAAGCATAGGTTTTTCCTGT 8479
```

RESULT 8
US-10-041-018-19
; Sequence 19, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 12753
; TYPE: DNA
; ORGANISM: Pantoea agglomerans
US-10-041-018-19

```
  Query Match             30.5%;  Score 2135.2;  DB 16;  Length 12753;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 3263;  Conservative    0;  Mismatches 1765;  Indels   39;  Gaps    4;

Qy          1479 CTGTTTCGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCT 1538
                 ||| | ||  |||| |  |   |  |  | |||   ||||||  |||| ||||   |  ||||
Db          5780 CTGCTGCGGCTGATTGCGGCCATGTCATCCCTGACCGATGTGCTCTGCCAGCAGTTGCCC 5839

Qy          1539 GCGGTACTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGC 1598
                 ||  | ||  |     ||    | || |   |||| | ||| || |||||  | || ||
Db          5840 GCTATTCTACAGCGGCTGGCGGTGGACGCGCTGATTGTCGATGAGATGGAGCCCGCCGGA 5899

Qy          1599 GGATTGGTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTC 1658
                  |  ||||||| || |||||||   || || ||| | ||  | ||||||||  ||||||||
Db          5900 AGCCTGGTCGCCGAGGCGCTGGGACTACCATTTATCTCTATTGCCTGCGCGCTGCCGGTC 5959

Qy          1659 AATCGTGAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAA 1718
                 || || ||  | ||  | ||||  |||||||||| || |  |  ||  |  |||| | | |
Db          5960 AACCGCGAGCCGGGTCTGCCGCTGCCGGTGATGCCGTTTCACTACGCCGAGGATAAGAGA 6019

Qy          1719 GCGCTGAAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGC 1778
                 || |||   ||||||||| ||||      ||||| |||    | |||   |  |||||
Db          6020 GCCCTGCGGCGTTTTCAGGTCAGCGAACGGATCTACGATGCGCTGATGTACCCGCACGGG 6079

Qy          1779 GACGTGATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAG 1838
                  |   |||||||    ||||| | |   |||  |||| |||||||  |  |  |  |  ||
Db          6080 CAGACGATCCTGCGCCACGCCCAGCGCTTTGGTTTGCCGGAGCGCAGGCGTCTCGACGAG 6139

Qy          1839 TGCCTGTCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAG 1898
                 || || |||||||||||| || || ||||||   || |||||| | || || ||||| | |
Db          6140 TGTCTCTCGCCGCTGGCGCAGATTAGCCAGTCCGTTCCGGCCCTCGACTTCCCACGCCGG 6199

Qy          1899 CAACTGCCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCG 1958
                    ||||    ||| | ||||  |||||| |||| |||         | ||   |||||
Db          6200 GCGCTGCCGAACTGTTTTCACTACGTGGGAGCACTGCGC---------TATCAGCCGCCG 6250

Qy          1959 CTCCATGCGCCCTGGCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTG 2018
                 |  || |      |  |  | |   |   ||||  | | | ||||||||||||| || || ||
Db          6251 CCGCAGGTAGAACGCTCGCCACGCAGCACGCCGCGGATCTTTGCCTCGCTGGGCACCCTC 6310
```

```
Qy   2019 CAAGGCCATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTA 2078
Db   6311 CAGGGCCACCGTCTACGCCTGTTTCAGAAGATCGCCCGCGCCTGTGCCAGCGTGGGGGCG 6370
Qy   2079 TCGCTGGTGATCGCCCATTGTGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTC 2138
Db   6371 GAGGTGACCATTGCCCACTGCGATGGCCTGACGCCCGCCCAGGCCGACTCGCTCTACGCC 6430
Qy   2139 GCTGGCGCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAG 2198
Db   6431 TGCGGCGCGACGGAGGTGGTCAGCTTTGTCGACCAGCCGCGCTACGTTGCCGAGGCTAAT 6490
Qy   2199 CTGTTTATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCG 2258
Db   6491 CTGGTGATCACCCACGGCGGTCTCAATACCGTACTGGATGCGCTGGCTGCCGCGACGCCG 6550
Qy   2259 ATGCTGGCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCAT 2318
Db   6551 GTGCTGGCGGTGCCACTCTCTTTCGACCAGCCCGCCGTGGCTGCCCGGCTGGTCTATAAC 6610
Qy   2319 GACGTTGGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAA 2378
Db   6611 GGGCTGGGTCGCCGGGTATCGCGCTTTGCCAGACAGCAGACGCTGGCGGATGAGATTGCC 6670
Qy   2379 CAGCTGCTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAG 2438
Db   6671 CAACTGCTGGGGGATGAGACGCTGCATCAGCGTCTGGCGACGGCCCGCCAGCAGCTTAAC 6730
Qy   2439 CGCGCAGGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAA 2498
Db   6731 GACGCCGGGGGCACGCCCCGTGCGGCGACCCTGATTGAACAGGCCAT------------- 6777
Qy   2499 GTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTG 2558
Db   6778 -----------AGCAGGGAGTGAGAGCGTATCGTGAGGGATCTGATTTTAGTCGGCGGCG 6826
Qy   2559 GACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCC 2618
Db   6827 GCCTGGCCAACGGGCTGATCGCCTGGCGTCTGCGCCAGCGCTACCCGCAGCTTAACCTGC 6886
Qy   2619 TGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCG 2678
Db   6887 TGCTGATCGAGGCCGGGGAGCAGCCCGGCGGGAACCATACCTGGTCATTCCATGAAGACG 6946
Qy   2679 ATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTT 2738
Db   6947 ATCTGACTCCCGGGCAGCACGCCTGGCTGGCCCCGCTGGTGGCCCACGCCTGGCCGGGCT 7006
Qy   2739 ATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTTCCATCGCAT 2798
Db   7007 ATGAGGTGCAGTTTCCCGATCTTCGCCGTCGCCTCGCGCGCGGCTACTACTCCATTACCT 7066
Qy   2799 CAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGACAAACACAG 2858
Db   7067 CAGAGCGCTTTGCCGAGGCCCTGCATCAGGCGCTGGGGGAGAACATCTGGCTAAACTGTT 7126
Qy   2859 CCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGC 2918
Db   7127 CGGTGAGCGAGGTGTTACCCAATAGCGTGCGCCTTGCCAACGGTGAGGCGCTGCTTGCCG 7186
Qy   2919 AAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGG 2978
Db   7187 GAGCGGTGATTGACGGACGCGGCGTGACCGCCAGTTCGGCGATGCAAACCGGCTATCAGC 7246
Qy   2979 TGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGA 3038
Db   7247 TCTTTCTTGGTCAGCAGTGGCGGCTGACACAGCCCCACGGCCTGACCGTACCGATCCTGA 7306
Qy   3039 TGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCG 3098
Db   7307 TGGATGCCACGGTGGCGCAGCAGCAGGGCTATCGCTTTGTCTACACGCTGCCGCTCTCCG 7366
Qy   3099 CCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACA 3158
```

```
Db   7367 CCGACACGCTGCTGATCGAGGATACGCGCTACGCCAATGTCCCGCAGCGTGATGATAATG 7426
Qy   3159 CCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGC 3218
Db   7427 CCCTACGCCAGACGGTTACCGACTATGCTCACAGCAAAGGGTGGCAGCTGGCCCAGCTTG 7486
Qy   3219 TGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATCGATTCTGGC 3278
Db   7487 AACGCGAGGAGACCGGCTGTCTGCCGATTACCCTGGCGGGTGACATCCAGGCTCTGTGGG 7546
Qy   3279 AACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCG 3338
Db   7547 CCGATGCGCCGGGCGTGCCGCGCTCGGGAATGCGGGCTGGGCTATTTCACCCTACCACTG 7606
Qy   3339 GTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCG 3398
Db   7607 GCTATTCGCTGCCGCTGGCGGTGGCCCTTGCCGACGCGATTGCCGACAGCCCGCGGCTGG 7666
Qy   3399 ATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGC 3458
Db   7667 GCAGCGTTCCGCTCTATCAGCTCACCCGGCAGTTTGCCGAACGCCACTGGCGCAGGCAGG 7726
Qy   3459 GATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGC 3518
Db   7727 GATTCTTCCGCCTGCTGAACCGGATGCTTTTCCTGGCCGGGCGCGAGGAGAACCGCTGGC 7786
Qy   3519 GCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGC 3578
Db   7787 GGGTGATGCAGCGCTTTTATGGGCTGCCGGAGCCCACCGTAGAGCGCTTTTACGCCGGTC 7846
Qy   3579 AACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTG 3638
Db   7847 GGCTCTCTCTCTTTGATAAGGCCCGCATTTTGACGGGCAAGCCACCGGTTCCGCTGGGCG 7906
Qy   3639 AAGCGCTGCGCGCGCTGTTGA---ATTCTGTCGAACCAGGGAAGAAAAAATGAAACGCAC 3695
Db   7907 AAGCCTGGCGGGCGGCGCTGAACCATTTTCCTGACAGACGAGATAAAGGATGAAAAAAAC 7966
Qy   3696 TTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGCGGCGGG 3755
Db   7967 CGTTGTGATTGGCGCAGGCTTTGGTGGCCTGGCGCTGGCGATTCGCCTGCAGGCGGCAGG 8026
Qy   3756 CATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGTGTTTGA 3815
Db   8027 GATCCCAACCGTACTGCTGGAGCAGCGGGACAAGCCCGGCGGTCGGGCCTACGTCTGGCA 8086
Qy   3816 GGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGCCATCGA 3875
Db   8087 TGACCAGGGCTTTACCTTTGACGCCGGGCCGACGGTGATCACCGATCCTACCGCGCTTGA 8146
Qy   3876 AGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCCGGTAAC 3935
Db   8147 GGCGCTGTTCACCCTGGCCGGCAGGCGCATGGAGGATTACGTCAGGCTGCTGCCGGTAAA 8206
Qy   3936 GCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAATCAGCC 3995
Db   8207 ACCCTTCTACCGACTCTGCTGGGAGTCCGGGAAGACCCTCGACTATGCTAACGACAGCGC 8266
Qy   3996 GCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCGTCAATT 4055
Db   8267 CGAGCTTGAGGCGCAGATTACCCAGTTCAACCCCCGCGACGTCGAGGGCTACCGGCGCTT 8326
Qy   4056 TCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGTGCCGTT 4115
Db   8327 TCTGGCTTACTCCCAGGCGGTATTCCAGGAGGGATATTTGCGCCTCGGCAGCGTGCCGTT 8386
Qy   4116 TCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGCATGGCG 4175
Db   8387 CCTCTCTTTTCGCGACATGCTGCGCGCCGGGCCGCAGCTGCTTAAGCTCCAGGCGTGGCA 8446
Qy   4176 CAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGCGTTTTC 4235
```

```
Db          8447 GAGCGTCTACCAGTCGGTTTCGCGCTTTATTGAGGATGAGCATCTGCGGCAGGCCTTCTC 8506

Qy          4236 CTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATACCTTAAT 4295
                  ||||||||  ||||||| ||||| || || ||  | || || || ||||| ||| | ||
Db          8507 GTTCCACTCCCTGCTGGTAGGCGGCAACCCCTTCACCACCTCGTCCATCTACACCCTGAT 8566

Qy          4296 TCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGCGCTGGT 4355
                  || || || ||||| || ||||| || ||||| ||    ||||||||||| ||||||||
Db          8567 CCACGCCCTTGAGCGGGAGTGGGGGGTCTGGTTCCCTGAGGGCGGCACCGGGGCGCTGGT 8626

Qy          4356 GCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGCCGAAGT 4415
                 | | ||||||| |   |||||    ||  ||||||| ||| |   ||| || |||   ||
Db          8627 GAACGGCATGGTGAAGCTGTTTACCGATCTGGGCGGGGAGATCGAACTCAACGCCCGGGT 8686

Qy          4416 GAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGGACGACG 4475
                      ||||||       ||  || ||| | |||   || | | | | ||  || ||
Db          8687 CGAAGAGCTGGTGGTGGCCGATAACCGCGTAAGCCAGGTCCGGCTGGCGGATGGTCGGAT 8746

Qy          4476 CTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAAACTGCT 4535
                 ||| ||  ||| ||| || ||||| || || |||||||||| | |||||  | || |||||
Db          8747 CTTTGACACCGACGCCGTAGCCTCGAACGCTGACGTGGTGAACACCTATAAAAAGCTGCT 8806

Qy          4536 TCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCATGAGCAA 4595
                   ||||||||||| |||    ||| || ||| || ||||| |||| ||| ||||||||||
Db          8807 CGGCCACCATCCGGTGGGGCAGAAGCGGGCGGCAGCGCTGGAGCGCAAGAGCATGAGCAA 8866

Qy          4596 CTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCACCACAC 4655
                 |||||||||||| ||||| || ||||||||| ||||| |||   ||||| |||||||| ||
Db          8867 CTCGCTGTTTGTGCTCTACTTCGGCCTGAACCAGCCTCATTCCCAGCTGGCGCACCATAC 8926

Qy          4656 CGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAGCCAGCT 4715
                 | ||||||||||| || || || || ||| ||||||| ||||| || | | |||||  |||
Db          8927 CATCTGTTTTGGTCCCCGCTACCGGGAGCTGATCGACGAGATCTTTACCGGCAGCGCGCT 8986

Qy          4716 GGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCTGGCACC 4775
                 ||| || || || || || ||||||||| ||||||||||   | |||||| ||||| || ||
Db          8987 GGCGGATGACTTCTCGCTCTACCTGCACTCGCCCTGCGTGACCGATCCCTCGCTCGCGCC 9046

Qy          4776 GCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGCTGACAT 4835
                  ||||||||||| |||||| || ||| | || |||||||||||||| |||| |||     |
Db          9047 TCCCGGCTGCGCCAGCTTCTACGTGCTGGCCCCGGTGCCGCATCTTGGCAACGCGCCGCT 9106

Qy          4836 CGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGAGCAGCA 4895
                  ||||||   ||||| || |||    ||||||| || || ||||  || || ||  |||
Db          9107 GGACTGGGCGCAGGAGGGGCCGAAGCTGCGCGACCGCATCTTTGACTACCTTGAAGAGCG 9166

Qy          4896 CTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTTTGATTT 4955
                 ||| ||||| ||  | |||   ||    ||||| ||  | || ||||| |||   || ||
Db          9167 CTATATGCCCGGCCTGCGTAGCCAGCTGGTGACCCAGCGGATCTTTACCCCGGCAGACTT 9226

Qy          4956 TCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTTGACGCA 5015
                  | |||||||||| |||| ||||  || ||||| || ||  | ||||||| |  |||| ||
Db          9227 CCACGACACGCTGGATGCGCATCTGGGATCGGCCTTCTCCATCGAGCCGCTGCTGACCCA 9286

Qy          5016 AAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCTGGTGGG 5075
                 |||||||||||||||||||||| ||||||||   ||| ||   ||| ||||| ||||||||
Db          9287 AAGCGCCTGGTTCCGCCCGCACAACCGCGACAGCGACATTGCCAACCTCTACCTGGTGGG 9346

Qy          5076 TGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGCCACCGC 5135
                  || ||||| || || || |||||| | || |||||  | |  ||||| || ||||||||
Db          9347 CGCAGGTACTCACCCTGGGGCGGGCATTCCTGGCGTAGTGGCCTCGGCGAAAGCCACCGC 9406

Qy          5136 CAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAAGTAACG 5195
                 ||| |||||| | |||||||      |||||  | || |||  | ||||| || |  |||
Db          9407 CAGCCTGATGATTGAGGATC---TGCAATGAGCCAACCGCCGCTGCTTGACCACGCCACG 9463

Qy          5196 CAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGATGCACCG 5255
                 || ||||||||    ||||||||| ||||| |||||||| || |||||||| ||  |  |
Db          9464 CAGACCATGGCCAACGGCTCGAAAAGTTTTGCCACCGCTGCGAAGCTGTTCGACCCGGCC 9523

Qy          5256 ACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGAT 5315
                 || ||||| |||  |||||||||| ||  | ||||| || ||||||||||| || |||||
Db          9524 ACCCGCCGTAGCGTGCTGATGCTCTACACCTGGTGCCGCCACTGCGATGACGTCATTGAC 9583

Qy          5316 GGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATG 5375
                 |  || || |  |||   | | ||  |      || |  || || ||       ||  ||
Db          9584 GACCAGACCCACGGCTTCGCCAGCGAGGCCGCGGCGGAGGAGGAGGCCACCCAGCGCCTG 9643

Qy          5376 CAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGCG 5435
                    |  ||||  |     ||||     || |     || ||  | ||| | || ||||| 
Db          9644 GCCCGGCTGCGCACGCTGACCCTGGCGGCGTTTGAAGGGGCCGAGATGCAGGATCCGGCC 9703

Qy          5436 TTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGAT 5495
                 ||    || |||||||| ||||| | |  |||    |  |||  |   ||||| | |||
Db          9704 TTCGCTGCCTTTCAGGAGGTGGCGCTGACCCACGGTATTACGCCCCGCATGGCGCTCGAT 9763

Qy          5496 CATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACACG 5555
                 || || || ||||| || ||||| |      |    |  || |  | ||| || || |||
Db          9764 CACCTCGACGGCTTTGCGATGGACGTGGCTCAGACCCGCTATGTCACCTTTGAGGATACG 9823

Qy          5556 CTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGGC 5615
                 ||||| |||||||||||||| ||||||||||| ||| ||||||||||  | || ||||||
Db          9824 CTGCGCTACTGCTATCACGTGGCGGGCGTGGTGGGTCTGATGATGGCCAGGGTGATGGGC 9883

Qy          5616 GTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCACT 5675
                 ||||| || ||   |||||| |||| |||||||||| | || ||||| |||||||| || 
Db          9884 GTGCGGGATGAGCGGGTGCTGGATCGCGCCTGCGATCTGGGGCTGGCCTTCCAGCTGACG 9943

Qy          5676 AACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATCC 5735
                 || || || || || ||| | || ||||| |  | ||  |||||||||||||||
Db          9944 AATATCGCCCGGGATATTATTGACGATGCGGCTATTGACCGCTGCTATCTGCCCGCCGAG 10003

Qy          5736 TGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGCG 5795
                 |||||  |  | || ||  |   | | || |      | || | | |  |||| || |||
Db         10004 TGGCTGCAGGATGCCGGGCTGACCCCGGAGAACTATGCCGCGCGGGAGAATCGGGCCGCG 10063

Qy          5796 CTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCGA 5855
                 || ||     |||| | |||  |  | |  |  || || || || ||   |||  | |
Db         10064 CTGGCGCGGGTGGCGGAGCGGCTTATTGATGCCGCAGAGCCGTACTACATCTCCTCCCAG 10123

Qy          5856 TCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGCGGCGTT 5915
                  ||||  |||  | | | |||| |||||  || ||||| ||||| || || ||| ||||
Db         10124 GCCGGGCTACACGATCTGCCGCCGCGCTGCGCCTGGGCGATCGCCACCGCCCGCAGCGTC 10183

Qy          5916 TATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCACGGCAG 5975
                 || || || || ||  | || ||  |    || || |    ||||||||||   || |||
Db         10184 TACCGGGAGATCGGTATTAAGGTAAAAGCGGCGGGAGGCAGCGCCTGGGATCGCCGCCAG 10243

Qy          5976 CGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCGATCACT 6035
                 | |||||| ||||||||||||| | ||  ||||| ||   |   | ||   ||  ||
Db         10244 CACACCAGCAAAGGTGAAAAAATTGCCATGCTGATGGCGGCACCGGGGCAGGTTATTCGG 10303

Qy          6036 TCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTCGTTGATTT 6095
                  ||     | |  |       ||||||||||| ||||| |||||||||||   ||
Db         10304 GCGAAGACGACGAGGGTGACGCCGCGTCCGGCCGGTCTTTGGCAGCGTCCCGTTTAGGCG 10363

Qy          6096 TACGTCCGTGACGCTGGCGCAGCGTGGCTTGCAGCTTATTCAGCGGTGGCGCGTAGAGGA 6155
                   || || ||||| |  |||||  | || || || |      ||    | ||||| |  |
Db         10364 GGCGGCCATGACGTTCACGCAGGATCGCCTGTAGGTCGGCAGGCTTGCGGGCGTAAATAA 10423

Qy          6156 AACCAAACGACACGCAGCCTTCACGCCCGCGCACCGCATGATGCATGCGGTGCGCCATGT 6215
                 |||| || || |||||||| || || ||||||||||| || |||| |||||| || | ||
Db         10424 AACCGAAGGAGACGCAGCCCTCCCGGCCGCGCACCGCGTGGTGCAGGCGGTGGGCGACGT 10483

Qy          6216 ATAAGCGCTTAAGATAGCCTTTGCGCGGGATATAGCGGAACGGCCAGCGTTGATGCACCA 6275
                 | |  ||||| || |||||   |||||||||  || |||| |||||||| ||||||||||
Db         10484 AGAGCCGCTTCAGGTAGCCCCGGCGCGGGATCCAGTGGAAGGGCCAGCGCTGATGCACCA 10543

Qy          6276 GGCCATCGTGCACCATGAAGTAGAGCGCGCCGTACGTCGTCATTCCGGCACCAATCCACT 6335
                 | || |||||||||| |||||||||||  ||| ||   |||||| |||   ||||||||||
Db         10544 GACCGTCGTGCACCAGGAAGTAGAGCAGGCCATAGACCGTCATGCCGCAGCCAATCCACT 10603

Qy          6336 GCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGTACCGCAAACA 6395
                 |||| ||||| | |||   |   || ||   ||||||| | || ||||   | |||||||
Db         10604 GCAGGGGCCAAACGCCCGCCGTGCCCACGGCAATCAGCGCGATAGCCACCCCGGCAAACA 10663
```

```
Qy       6396 CCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGGTGCGACAGAT 6455
              |||||||| | |||||||| ||||||||    ||  || ||||   |||||| |||  ||
Db      10664 CCACCGCAAAGAGATCGTTTAGCTCAAATACGCCCTTGCGCGGGGTATGGTGTGACTCAT 10723

Qy       6456 GCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCTACGATTTCCA 6515
              |||||| ||||||||| |||||||| |||||    || |  | ||| || | |   ||||
Db      10724 GCCAGCGCCATCCCCAGCCGTGCATAATGTAGCGGTGGGTAAACGCGGCGATGCCCTCCA 10783

Qy       6516 TCACCACCACGGTTGCCAACAAGATAA 6542
              || | |  ||| |    |  | ||| |
Db      10784 TCGCAATAACGCTCAAGATGACGATTA 10810


RESULT 9
US-10-474-536-45
; Sequence 45, Application US/10474536
; Publication No. US20040176570A1
; GENERAL INFORMATION:
;  APPLICANT: Bacher, Adelbert
;  APPLICANT:  Rohdich, Felix
;  APPLICANT:  Adam, Petra
;  APPLICANT:  Amslinger, Sabine
;  APPLICANT:  Eisenreich, Wolfgang
;  APPLICANT:  Hecht, Stefan
;  TITLE OF INVENTION: INTERMEDIATES AND ENZYMES OF THE NON-MEVALONATE ISOPRENOID
;  TITLE OF INVENTION:  PATHWAY
;  FILE REFERENCE: 9286.24
;  CURRENT APPLICATION NUMBER: US/10/474,536
;  CURRENT FILING DATE:  2003-10-10
;  PRIOR APPLICATION NUMBER: PCT/EP02/04005
;  PRIOR FILING DATE: 2002-04-10
;  PRIOR APPLICATION NUMBER: DE 102 01 458.2
;  PRIOR FILING DATE: 2002-01-16
;  PRIOR APPLICATION NUMBER: DE 101 55 084.7
;  PRIOR FILING DATE: 2001-11-09
;  PRIOR APPLICATION NUMBER: DE 101 30 236.3
;  PRIOR FILING DATE: 2001-06-22
;  PRIOR APPLICATION NUMBER: DE 101 18 166.3
;  PRIOR FILING DATE: 2001-04-11
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
;   LENGTH: 7494
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pBScaro14
US-10-474-536-45

  Query Match             23.1%;  Score 1618.8;  DB 17;  Length 7494;
  Best Local Similarity   66.0%;  Pred. No. 0;
  Matches 2375;  Conservative    0;  Mismatches 1217;  Indels    6;  Gaps    2;

Qy       2490 CAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGG 2549
              || | |  || |   ||   |  || |     ||||   |||| || ||| ||||| | |
Db       3213 CACCGGTCAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCG 3272

Qy       2550 TCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAAC 2609
              | || |||||||| ||||||||| | || || |||||||| |  ||   ||| ||  |
Db       3273 TGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATA 3332

Qy       2610 TGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTC 2669
              ||       || |  | ||   ||   | ||  |||  || |||||||| ||||| ||||
Db       3333 TGCGTATTTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTC 3392

Qy       2670 ATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTT 2729
              | |||   ||| | |  |     |||| ||| ||| |    |||||| |       |  |
Db       3393 ACCACGATGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACT 3452

Qy       2730 GGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTT 2789
              || | |  |||||||| || |||||  | |  || || || ||| || | |  || | ||
Db       3453 GGCCCGACTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTT 3512
```

```
Qy       2790 CCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGA 2849
                ||  | ||      ||| ||      | |  |      | | ||     |  ||||||
Db       3513 GTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGA 3572

Qy       2850 CAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAAC 2909
                 | || || ||   | |||| ||  |       ||   | ||    | || |  |
Db       3573 TGGATACCGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTA 3632

Qy       2910 TTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGG 2969
              | | ||| |  | ||||||||| || || ||      | | |   ||   |||    |||
Db       3633 TCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGG 3692

Qy       2970 GTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGC 3029
              | |  |||| |||| |||| || || ||||   ||   || ||||| ||  |   |   |
Db       3693 GCTTCCAGGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTC 3752

Qy       3030 CGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGC 3089
              | ||  | ||||||||||| ||||||||||||   |||||||| || || ||||  ||||
Db       3753 CCATTATCATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGC 3812

Qy       3090 CGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGG 3149
              |||||    |     |  | || ||||||||| || || ||  || |  |  |  |  | |
Db       3813 CGCTCTCGCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAG 3872

Qy       3150 CGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGA 3209
                    |   ||| || ||  | ||   ||||||||||   || || || |||  |||
Db       3873 ATCCTGAATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTC 3932

Qy       3210 GTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATC 3269
                || |||||||| ||||| || |||   ||||| ||||| |||   |||||   |||
Db       3933 AGACACTGCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACG 3992

Qy       3270 GATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATG 3329
               |||||||   |||||||||  ||  || || ||  | || ||||| ||||| |||
Db       3993 CATTCTGG---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATC 4049

Qy       3330 CCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGT 3389
              | |||||||| || ||  ||||   || || ||  | || ||     |    || ||
Db       4050 CTACCACCGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTG 4109

Qy       3390 TGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGC 3449
                  |       |   | |  |   |||    ||    |  |||||||| |||  ||||
Db       4110 ATGTCTTTACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGC 4169

Qy       3450 GCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGC 3509
                 | ||| | ||||||||  |||| | ||||||||||||||| |||||     |  |
Db       4170 AGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATT 4229

Qy       3510 AGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTT 3569
                |||||||| || ||||| ||||||||  |  |   ||  |  ||||||  ||| ||||
Db       4230 CACGCTGGCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTT 4289

Qy       3570 ACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGC 3629
              | || ||  ||||    |||  ||||      || ||||||| ||||||||||||| || |
Db       4290 ATGCGGGAAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTC 4349

Qy       3630 CCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAA 3689
              |  |    | |||  |||  ||  |  |||   ||        | |   ||  | |||||
Db       4350 CGGTATTAGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAA 4409

Qy       3690 ACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGC 3749
              ||  |||   || ||||| |||||||| || |||||||| ||||| ||||| || |||||
Db       4410 ACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGC 4469

Qy       3750 GGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGT 3809
               ||||| || ||   |    | || || || || ||||| ||||| || || || |||||
Db       4470 TGCGGGGATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGT 4529

Qy       3810 GTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGC 3869
               |  |||||    || |||||||| ||||| || || ||||| |||||||||||||| ||
Db.      4530 CTACGAGGATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGC 4589

Qy       3870 CATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCC 3929
```

```
          ||| |||||  ||||  | ||||||||||||  | | |  || || ||||| ||| ||||
Db   4590 CATTGAAGAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCC 4649

Qy   3930 GGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAA 3989
          ||| ||||| || || |||||||| |||||    || ||     || ||||||| ||  |
Db   4650 GGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGA 4709

Qy   3990 TCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCG 4049
          |||  | | ||| ||   ||||||      ||| ||||| |  ||||| ||||| |||||
Db   4710 TCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCG 4769

Qy   4050 TCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGT 4109
          ||| ||||| | ||||||||| |  || |||| ||| || ||||| || ||||| || ||
Db   4770 TCAGTTTCTGGACTATTCACGCGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGT 4829

Qy   4110 GCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGC 4169
           || ||| |   | |  | |||||||| ||| ||| || ||  |||     ||||| ||
Db   4830 CCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGC 4889

Qy   4170 ATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGC 4229
          |||| | |||||| ||||| | ||| || | |  ||  | || || ||||||||| |||||
Db   4890 ATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGC 4949

Qy   4230 GTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATAC 4289
          |||||| ||||||||  || |||||||||| ||||| || || || |||| || |||||
Db   4950 GTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATAC 5009

Qy   4290 CTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGC 4349
           || || || |||||||||||||| ||||||||| |||||||||||| ||||||||||||||
Db   5010 GTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGC 5069

Qy   4350 GCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGC 4409
            | || ||||| |||      |||||  ||||  |||| |||||  |  |  | || ||
Db   5070 ATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGC 5129

Qy   4410 CGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGG 4469
          |  ||| |||||  ||||||| |  || ||    |||   | ||| || ||||||| |||
Db   5130 CAGAGTCAGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGG 5189

Qy   4470 ACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAA 4529
           ||  | |||    |    || || || || ||||| || ||||| ||||||||   | |
Db   5190 TCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGA 5249

Qy   4530 ACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCAT 4589
           ||| |  |||| || ||    ||  | || |   | |     ||| ||  |||||||||
Db   5250 CCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCAT 5309

Qy   4590 GAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCA 4649
          ||| ||||| |||||||| |||||||||||  ||||||| |  ||||| ||||||||||||
Db   5310 GAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCA 5369

Qy   4650 CCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAG 4709
           ||||| || ||||| ||||||||||| ||  ||| |||| || || ||||| ||
Db   5370 TCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGA 5429

Qy   4710 CCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCT 4769
              || ||||| || || ||||||||| ||||||||||||||   ||  ||| |||| ||
Db   5430 TGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACT 5489

Qy   4770 GGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGC 4829
          ||| ||    || ||||||||| |  |||||||| ||||||||||||||||| | ||||||||
Db   5490 GGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGC 5549

Qy   4830 TGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGA 4889
            || |||||||||      || || ||     | ||||| || ||||||||| || || ||
Db   5550 GAACCTCGACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGA 5609

Qy   4890 GCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTT 4949
          |||||| |||||||| || |||||    ||  | || || ||| | ||||||||||||||
Db   5610 GCAGCATTACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTT 5669

Qy   4950 TGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTT 5009
          |||||||||||||  |||  ||||| |||| ||||| || |||||  |||||||  || |
```

```
Db   5670 TGATTTTCGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCT 5729

Qy   5010 GACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCT 5069
           || || |||||||||| || |||||||||||||||||      || |  |||||||| ||
Db   5730 TACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCT 5789

Qy   5070 GGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGC 5129
          ||| || || || ||||||| ||||| ||| | || ||||| ||||| ||||| || ||
Db   5790 GGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGC 5849

Qy   5130 CACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAA 5189
           || ||  |  |||||||||||||||        |||||   | | | |||||  | ||
Db   5850 GACAGCAGGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCAT 5906

Qy   5190 GTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGAT 5249
          |      |||| ||||| || |||||||| ||||| || || ||| | ||| | | ||||||
Db   5907 GCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGAT 5966

Qy   5250 GCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTG 5309
          |||   || || ||||||   |||||||| || || |||| || || || || |||||
Db   5967 GCAAAAACCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTT 6026

Qy   5310 ATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCA 5369
          ||||| |  || |||||||||      | |  ||||| |||| |  ||    |  |   |
Db   6027 ATTGACGATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAA 6086

Qy   5370 CGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAA 5429
          ||| ||  ||| ||  | ||  |||| ||||  |||||    ||  |||| ||| | |||
Db   6087 CGTCTGATGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAA 6146

Qy   5430 CCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCG 5489
          |||||||||  ||| ||||||||||||||| ||   |||  |  |  | |        |||
Db   6147 CCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCG 6206

Qy   5490 TTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGAT 5549
          ||||||||||||||||||||||| |||||||| |||| | |  || ||        | |||
Db   6207 TTTGATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGAT 6266

Qy   5550 GACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTA 5609
          || ||||||||| || |||||||||| || ||||| |||||| ||||||||||||||   |
Db   6267 GATACGCTGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATC 6326

Qy   5610 ATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAG 5669
          ||||||||||| ||  | ||   ||| || | |||||| ||  | || ||||| || |||
Db   6327 ATGGGCGTGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAG 6386

Qy   5670 CTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCG 5729
           | || || ||||| ||||| ||||| || |||||  |    || ||||| |||||||||
Db   6387 TTGACCAATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCG 6446

Qy   5730 CAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGT 5789
            |  |||||| || || |  ||  |   |   || |      | |||||  || | |||
Db   6447 GCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGT 6506

Qy   5790 GCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCG 5849
             |||||   |    | ||    ||||| |||  ||| || ||||| ||  |||   ||
Db   6507 CAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCT 6566

Qy   5850 GCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGC 5909
          ||   | ||||  |  | || || || ||||| || || ||||| ||||||||||||
Db   6567 GCCACAGCCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAG 6626

Qy   5910 GGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCA 5969
             ||||| ||  |||| || ||||||||| | || ||||||  |||  ||||||||||
Db   6627 CAGGTTTACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAG 6686

Qy   5970 CGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCG 6029
          ||||||   || |  |     |||||| |  |||||||| ||   |   | |||  |||
Db   6687 CGGCAGTCAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCC 6746

Qy   6030 ATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTC 6087
           | ||||| ||  ||     || |   || || || ||    ||| ||||||||| || |
Db   6747 CTTACTTCCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC 6804
```

```
RESULT 10
US-10-474-536-46
; Sequence 46, Application US/10474536
; Publication No. US20040176570A1
; GENERAL INFORMATION:
;  APPLICANT: Bacher, Adelbert
;  APPLICANT:  Rohdich, Felix
;  APPLICANT:  Adam, Petra
;  APPLICANT:  Amslinger, Sabine
;  APPLICANT:  Eisenreich, Wolfgang
;  APPLICANT:  Hecht, Stefan
;  TITLE OF INVENTION: INTERMEDIATES AND ENZYMES OF THE NON-MEVALONATE ISOPRENOID
;  TITLE OF INVENTION:  PATHWAY
;  FILE REFERENCE: 9286.24
;  CURRENT APPLICATION NUMBER: US/10/474,536
;  CURRENT FILING DATE:  2003-10-10
;  PRIOR APPLICATION NUMBER: PCT/EP02/04005
;  PRIOR FILING DATE: 2002-04-10
;  PRIOR APPLICATION NUMBER: DE 102 01 458.2
;  PRIOR FILING DATE: 2002-01-16
;  PRIOR APPLICATION NUMBER: DE 101 55 084.7
;  PRIOR FILING DATE: 2001-11-09
;  PRIOR APPLICATION NUMBER: DE 101 30 236.3
;  PRIOR FILING DATE: 2001-06-22
;  PRIOR APPLICATION NUMBER: DE 101 18 166.3
;  PRIOR FILING DATE: 2001-04-11
;  NUMBER OF SEQ ID NOS: 58
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 46
;   LENGTH: 8547
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: pACYCcaro14
US-10-474-536-46

  Query Match             23.1%;  Score 1618.8;  DB 17;  Length 8547;
  Best Local Similarity   66.0%;  Pred. No. 0;
  Matches 2375;  Conservative    0;  Mismatches 1217;  Indels    6;  Gaps    2;

Qy       2490 CAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATTTTGG 2549

Db       2845 CACCGGTCAGCCTGTCTTAAGTGGGAGCGGCTATGCAACCGCATTATGATCTGATTCTCG 2904

Qy       2550 TCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCACAAC 2609

Db       2905 TGGGGGCTGGACTCGCGAATGGCCTTATCGCCCTGCGTCTTCAGCAGCAGCAACCTGATA 2964

Qy       2610 TGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCGTTTC 2669

Db       2965 TGCGTATTTTGCTTATCGACGCCGCACCCCAGGCGGGCGGGAATCATACGTGGTCATTTC 3024

Qy       2670 ATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTGCGTT 2729

Db       3025 ACCACGATGATTTGACTGAGAGCCAACATCGTTGGATAGCTCCGCTGGTGGTTCATCACT 3084

Qy       2730 GGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTATTGTT 2789

Db       3085 GGCCCGACTATCAGGTACGCTTTCCCACACGCCGTCGTAAGCTGAACAGCGGCTACTTTT 3144

Qy       2790 CCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTGTGGA 2849

Db       3145 GTATTACTTCTCAGCGTTTCGCTGAGGTTTTACAGCGACAGTTTGGCCCGCACTTGTGGA 3204

Qy       2850 CAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGTGAAC 2909

Db       3205 TGGATACCGCGGTCGCAGAGGTTAATGCGGAATCTGTTCGGTTGAAAAAGGGTCAGGTTA 3264

Qy       2910 TTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAGCTGG 2969

Db       3265 TCGGTGCCCGCGCGGTGATTGACGGGCGGGGTTATGCGGCAAATTCAGCACTGAGCGTGG 3324

Qy       2970 GTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAGCAGC 3029

Db       3325 GCTTCCAGGCGTTTATTGGCCAGGAATGGCGATTGAGCCACCCGCATGGTTTATCGTCTC 3384

Qy       3030 CGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACGCTGC 3089

Db       3385 CCATTATCATGGATGCCACGGTCGATCAGCAAAATGGTTATCGCTTCGTGTACAGCCTGC 3444

Qy       3090 CGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCGCTGG 3149

Db       3445 CGCTCTCGCCGACCAGATTGTTAATTGAAGACACGCACTATATTGATAATGCGACATTAG 3504

Qy       3150 CGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACGCTGA 3209

Db       3505 ATCCTGAATGCGCGCGGCAAAATATTTGCGACTATGCCGCGCAACAGGGTTGGCAGCTTC 3564

Qy       3210 GTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATCGATC 3269

Db       3565 AGACACTGCTGCGAGAAGAACAGGGCGCCTTACCCATTACTCTGTCGGGCAATGCCGACG 3624

Qy       3270 GATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTTCATG 3329

Db       3625 CATTCTGG---CAGCAGCGCCCCCTGGCCTGTAGTGGATTACGTGCCGGTCTGTTCCATC 3681

Qy       3330 CCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCGCTGT 3389

Db       3682 CTACCACCGGCTATTCACTGCCGCTGGCGGTTGCCGTGGCCGACCGCCTGAGTGCACTTG 3741

Qy       3390 TGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAGTGGC 3449

Db       3742 ATGTCTTTACGTCGGCCTCAATTCACCATGCCATTACGCATTTTGCCCGCGAGCGCTGGC 3801

Qy       3450 GCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCGCAGC 3509

Db       3802 AGCAGCAGGGCTTTTTCCGCATGCTGAATCGCATGCTGTTTTTAGCCGGACCCGCCGATT 3861

Qy       3510 AGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGCTTTT 3569

Db       3862 CACGCTGGCGGGTTATGCAGCGTTTTTATGGTTTACCTGAAGATTTAATTGCCCGTTTTT 3921

Qy       3570 ACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCGGTGC 3629

Db       3922 ATGCGGGAAAACTCACGCTGACCGATCGGCTACGTATTCTGAGCGGCAAGCCGCCTGTTC 3981

Qy       3630 CCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAATGAA 3689

Db       3982 CGGTATTAGCAGCATTGCAAGCCATTATGACGACTCATCGTTAAAGAGCGACTACATGAA 4041

Qy       3690 ACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGCAAGC 3749

Db       4042 ACCAACTACGGTAATTGGTGCAGGCTTCGGTGGCCTGGCACTGGCAATTCGTCTACAAGC 4101

Qy       3750 GGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCTATGT 3809

Db       4102 TGCGGGGATCCCCGTCTTACTGCTTGAACAACGTGATAAACCCGGCGGTCGGGCTTATGT 4161

Qy       3810 GTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCAGCGC 3869

Db       4162 CTACGAGGATCAGGGGTTTACCTTTGATGCAGGCCCGACGGTTATCACCGATCCCAGTGC 4221

Qy       3870 CATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGATGCC 3929

Db       4222 CATTGAAGAACTGTTTGCACTGGCAGGAAAACAGTTAAAAGAGTATGTCGAACTGCTGCC 4281

Qy       3930 GGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACAATAA 3989

Db       4282 GGTTACGCCGTTTTACCGCCTGTGTTGGGAGTCAGGGAAGGTCTTTAATTACGATAACGA 4341

Qy       3990 TCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCTATCG 4049

Db       4342 TCAAACCCGGCTCGAAGCGCAGATTCAGCAGTTTAATCCCCGCGATGTCGAAGGTTATCG 4401
```

```
Qy        4050 TCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCACGGT 4109
Db        4402 TCAGTTTCTGGACTATTCACGCGGTGTTTAAAGAAGGCTATCTAAAGCTCGGTACTGT 4461

Qy        4110 GCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGCAAGC 4169
Db        4462 CCCTTTTTTATCGTTCAGAGACATGCTTCGCGCCGCACCTCAACTGGCGAAACTGCAGGC 4521

Qy        4170 ATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTCAGGC 4229
Db        4522 ATGGAGAAGCGTTTACAGTAAGGTTGCCAGTTACATCGAAGATGAACATCTGCGCCAGGC 4581

Qy        4230 GTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCTATAC 4289
Db        4582 GTTTTCTTTCCACTCGCTGTTGGTGGGCGGCAATCCCTTCGCCACCTCATCCATTTATAC 4641

Qy        4290 CTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCGGCGC 4349
Db        4642 GTTGATACACGCGCTGGAGCGTGAGTGGGGCGTCTGGTTTCCGCGTGGCGGCACCGGCGC 4701

Qy        4350 GCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGAATGC 4409
Db        4702 ATTAGTTCAGGGGATGATAAAGCTGTTTCAGGATCTGGGTGGCGAAGTCGTGTTAAACGC 4761

Qy        4410 CGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGGGCGG 4469
Db        4762 CAGAGTCAGCCATATGGAAACGACAGGAAACAAGATTGAAGCCGTGCATTTAGAGGACGG 4821

Qy        4470 ACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACGACAA 4529
Db        4822 TCGCAGGTTCCTGACGCAAGCCGTCGCGTCAAATGCAGATGTGGTTCATACCTATCGCGA 4881

Qy        4530 ACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGCGCAT 4589
Db        4882 CCTGTTAAGCCAGCACCCTGCCGCGGTTAAGCAGTCCAACAAACTGCAGACTAAGCGCAT 4941

Qy        4590 GAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCGCGCA 4649
Db        4942 GAGTAACTCTCTGTTTGTGCTCTATTTTGGTTTGAATCACCATCATGATCAGCTCGCGCA 5001

Qy        4650 CCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACAGCAG 4709
Db        5002 TCACACGGTTTGTTTCGGCCCGCGTTACCGCGAGCTGATTGACGAAATTTTTAATCATGA 5061

Qy        4710 CCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGTCGCT 4769
Db        5062 TGGCCTCGCAGAGGACTTCTCACTTTATCTGCACGCGCCCTGTGTCACGGATTCGTCACT 5121

Qy        4770 GGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCACCGC 4829
Db        5122 GGCGCCTGAAGGTTGCGGCAGTTACTATGTGTTGGCGCCGGTGCCGCATTTAGGCACCGC 5181

Qy        4830 TGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATCTGGA 4889
Db        5182 GAACCTCGACTGGACGGTTGAGGGGCCAAAACTACGCGACCGTATTTTTGCGTACCTTGA 5241

Qy        4890 GCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGCCGTT 4949
Db        5242 GCAGCATTACATGCCTGGCTTACGGAGTCAGCTGGTCACGCACCGGATGTTTACGCCGTT 5301

Qy        4950 TGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGATTTT 5009
Db        5302 TGATTTTCGCGACCAGCTTAATGCCTATCATGGCTCAGCCTTTTCTGTGGAGCCCGTTCT 5361

Qy        5010 GACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCTATCT 5069
Db        5362 TACCCAGAGCGCCTGGTTTCGGCCGCATAACCGCGATAAAACCATTACTAATCTCTACCT 5421

Qy        5070 GGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCAAGGC 5129
Db        5422 GGTCGGCGCAGGCACGCATCCCGGCGCAGGCATTCCTGGCGTCATCGGCTCGGCAAAAGC 5481

Qy        5130 CACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGAGCAA 5189
Db        5482 GACAGCAGGTTTGATGCTGGAGGATC---TGATTTGAATAATCCGTCGTTACTCAATCAT 5538

Qy        5190 GTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTTTGAT 5249
Db        5539 GCGGTCGAAACGATGGCAGTTGGCTCGAAAAGTTTTGCGACAGCCTCAAAGTTATTTGAT 5598

Qy        5250 GCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTG 5309
Db        5599 GCAAAAACCCGGCGCAGCGTACTGATGCTCTACGCCTGGTGCCGCCATTGTGACGATGTT 5658

Qy        5310 ATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCA 5369
Db        5659 ATTGACGATCAGACGCTGGGCTTTCAGGCCCGGCAGCCTGCCTTACAAACGCCCGAACAA 5718

Qy        5370 CGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAA 5429
Db        5719 CGTCTGATGCAACTTGAGATGAAAACGCGCCAGGCCTATGCAGGATCGCAGATGCACGAA 5778

Qy        5430 CCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCG 5489
Db        5779 CCGGCGTTTGCGGCTTTTCAGGAAGTGGCTATGGCTCATGATATCGCCCCGGCTTACGCG 5838

Qy        5490 TTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGAT 5549
Db        5839 TTTGATCATCTGGAAGGCTTCGCCATGGATGTACGCGAAGCGCAATACAGCCAACTGGAT 5898

Qy        5550 GACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTA 5609
Db        5899 GATACGCTGCGCTATTGCTATCACGTTGCAGGCGTTGTCGGCTTGATGATGGCGCAAATC 5958

Qy        5610 ATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAG 5669
Db        5959 ATGGGCGTGCGGGATAACGCCACGCTGGACCGCGCCTGTGACCTTGGGCTGGCATTTCAG 6018

Qy        5670 CTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCG 5729
Db        6019 TTGACCAATATTGCTCGCGATATTGTGGACGATGCGCATGCGGGCCGCTGTTATCTGCCG 6078

Qy        5730 CAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGT 5789
Db        6079 GCAAGCTGGCTGGAGCATGAAGGTCTGAACAAAGAGAATTATGCGGCACCTGAAAACCGT 6138

Qy        5790 GCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCG 5849
Db        6139 CAGGCGCTGAGCCGTATCGCCCGTCGTTTGGTGCAGGAAGCAGAACCTTACTATTTGTCT 6198

Qy        5850 GCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTCGC 5909
Db        6199 GCCACAGCCGGCCTGGCAGGGTTGCCCCTGCGTTCCGCCTGGGCAATCGCTACGGCGAAG 6258

Qy        5910 GGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGATTCA 5969
Db        6259 CAGGTTTACCGGAAAATAGGTGTCAAAGTTGAACAGGCCGGTCAGCAAGCCTGGGATCAG 6318

Qy        5970 CGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTTGGCG 6029
Db        6319 CGGCAGTCAACGACCACGCCCGAAAAATTAACGCTGCTGCTGGCCGCCTCTGGTCAGGCC 6378

Qy        6030 ATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCCTC 6087
Db        6379 CTTACTTCCCGGATGCGGGCTCATCCTCCCCGCCCTGCGCATCTCTGGCAGCGCCCGC 6436

RESULT 11
US-10-808-979-7
; Sequence 7, Application US/10808979
; Publication No. US20040268439A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2360 US NA
;  CURRENT APPLICATION NUMBER: US/10/808,979
;  CURRENT FILING DATE:  2004-03-25
;  PRIOR APPLICATION NUMBER: US 60/471,904
;  PRIOR FILING DATE: 2003-05-20
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 1482
;   TYPE: DNA
;   ORGANISM: Enterobacteriaceae strain DC260
US-10-808-979-7

  Query Match             21.2%;  Score 1482;  DB 18;  Length 1482;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1482;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 60

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 120

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 180

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 240

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 300

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 360

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 420

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 480

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 540

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 600

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 660

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 720

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 780

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 840

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 900

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 960

Qy       4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 1020

Qy       4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 1080

Qy       4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 1140

Qy       4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 1200

Qy       4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 1260

Qy       4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 1320

Qy       5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 1380

Qy       5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 1440

Qy       5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
              ||||||||||||||||||||||||||||||||||||||||||
Db       1441 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 1482


RESULT 12
US-10-808-979-3
; Sequence 3, Application US/10808979
; Publication No. US20040268439A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2360 US NA
;  CURRENT APPLICATION NUMBER: US/10/808,979
;  CURRENT FILING DATE:  2004-03-25
;  PRIOR APPLICATION NUMBER: US 60/471,904
;  PRIOR FILING DATE: 2003-05-20
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
;   LENGTH: 1278
;   TYPE: DNA
;   ORGANISM: Enterobacteriaceae strain DC260
US-10-808-979-3

  Query Match             18.3%;  Score 1278;  DB 18;  Length 1278;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1278;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1245 GTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCG 1304
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GTGGGCCACTTTGCGGTGATCGCGCCGCCGCTCTACAGCCACTTTCACGCGTTGCAGGCG 60

Qy       1305 TTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCGATGCC 1364
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 TTAGCACAAACGCTGCTGGCGCGCGGCCATCGCATCACATTCATCCAGCAAGCCGATGCC 120

Qy       1365 CGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGCATCCT 1424
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CGCACTTTGCTTAGCGACGAACGCATCGATTTTGTTGCCGTCGGCCAACAGACGCATCCT 180

Qy       1425 GCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTT 1484
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GCCGGTTCGCTGGCGCCCGTGTTGCATCGGCTGGCCTCGCCGGGCGGCCTGTCGCTGTTT 240

Qy       1485 CGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTGCGGTA 1544
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGCGTGATCGACGATCTCGCGTCCTGCACCGATATGCTGTGCCGCGAACTGCCTGCGGTA 300

Qy       1545 CTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCGGATTG 1604
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CTGAAAGCATTGAACATCGATGGCGTGATCGCCGACGAAATGGAAGCGGCGGGCGGATTG 360

Qy       1605 GTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGT 1664
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GTCGCTGAAGCGCTGCATCTGCCGTTTGTTTCGGTGGCCTGCGCCTTGCCGGTCAATCGT 420

Qy       1665 GAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTG 1724
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 GAAGCCGGGATTCCGCTTGCGGTGATGCCCTTCCGTTTTGCACAGGATGACAAAGCGCTG 480

Qy       1725 AAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCGACGTG 1784
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACGTTTTCAGGCCAGCAGCGATATCTATGATCGCATCATGCGTCGTCACGGCGACGTG 540

Qy       1785 ATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGTGCCTG 1844
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATCCTCAAACACGCGCGGGCGTTTAATTTGACGGAGCGGCGCGGATTACATCAGTGCCTG 600

Qy       1845 TCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTG 1904
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TCGCCGCTGGCACAAATCAGCCAGATGGTGCCGGCCTTTGATTTTCCACGTCAGCAACTG 660

Qy       1905 CCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGCTCCAT 1964
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CCCGCCTGCTATCACGCCGTGGGGCCACTCCGCGCCCCGGTTTCTCCTGCGCCGCTCCAT 720

Qy       1965 GCGCCCTGGCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGC 2024
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GCGCCCTGGCCAGCGCTGCGTCAGCCGGTGGTTTATGCCTCGCTGGGTACGCTGCAAGGC 780

Qy       2025 CATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTG 2084
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CATCGCTTCCGGCTGTTTCTGCATCTGGCGCAGGCGTGCCGCCAGCTGCGGCTATCGCTG 840

Qy       2085 GTGATCGCCCATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGC 2144
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GTGATCGCCCATTGTGGGGGATTAAACGCCGAACAGACGCATCAGCTGGAGCTCGCTGGC 900

Qy       2145 GCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTT 2204
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GCGGCGTGGGTGACGGATTTCGTCGATCAGCGCGCAGCCCTACAGCACGCGCAGCTGTTT 960

Qy       2205 ATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTG 2264
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATCACTCATGCCGGGTTAAACAGCGCGCTGGAAGCACTGGAATGCGGTACGCCGATGCTG 1020

Qy       2265 GCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTT 2324
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GCGCTGCCGATTGCTTTTGATCAGCCCGGCGTGGCGGCGCGCATTGAGTGGCATGACGTT 1080

Qy       2325 GGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTG 2384
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GGTCGCCGCGCATCACGCTTTAGCCGTGTTCATCAACTGGAGCAGCATCTGCAACAGCTG 1140

Qy       2385 CTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCA 2444
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CTGACCGACGATCGTTACGCGCTACGGATGTCAGCGATTCAGGCGCAGCTGCAGCGCGCA 1200

Qy       2445 GGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTG 2504
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 GGCGGTTGCCAGCGTGCCGCCGACATCGTCGAGCAGGCGCTGTGCCAGCAGCAAGTCGTG 1260

Qy       2505 CTGGCGGAGGCGACCTGA 2522
              ||||||||||||||||||
Db       1261 CTGGCGGAGGCGACCTGA 1278


RESULT 13
US-10-808-979-5
; Sequence 5, Application US/10808979
; Publication No. US20040268439A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2360 US NA
;  CURRENT APPLICATION NUMBER: US/10/808,979
;  CURRENT FILING DATE:  2004-03-25
;  PRIOR APPLICATION NUMBER: US 60/471,904
;  PRIOR FILING DATE: 2003-05-20
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 5
;   LENGTH: 1167
;   TYPE: DNA
;   ORGANISM: Enterobacteriaceae strain DC260
US-10-808-979-5

  Query Match             16.7%;  Score 1167;  DB 18;  Length 1167;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1167;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2522 ATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCG 2581
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGCGCACGCAATACGATGTGATTTTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCG 60

Qy       2582 CTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCAT 2641
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CTGCGTCTGCGTCAATTGCAGCCACAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCAT 120

Qy       2642 CCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGC 2701
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 CCGGCAGGCAATCATACCTGGTCGTTTCATCACAGCGATCTCAGCGCCGAACAACTTCGC 180

Qy       2702 TGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTG 2761
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 TGGCTGCAACCGCTGATTACCGTGCGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTG 240

Qy       2762 CGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTT 2821
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 CGCCGCAATCTGGACGGGGATTATTGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTT 300

Qy       2822 TACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACG 2881
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 TACGCGGCGATGGGTGACGATCTGTGGACAAACACAGCCGTACAACAGGTAAAACCCACG 360

Qy       2882 CAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGC 2941
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGGTGACGCTGGCGGATGGCCGTGAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGC 420
```

```
Qy       2942 CTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAG 3001
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 CTGCAGCCGACGCCACATCTGCAGCTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAG 480

Qy       3002 CTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAA 3061
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 CTGGCGCAGCCGCACGGCCTGCAGCAGCCGATCCTGATGGATGCCACCGTCGATCAGCAA 540

Qy       3062 GCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGAT 3121
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 GCGGGTTATCGTTTTGTCTACACGCTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGAT 600

Qy       3122 ACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGAC 3181
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ACCCATTACGTTAACCAGCCCGCGCTGGCGGAGAACACCGCTCGTCAGCACATCGCCGAC 660

Qy       3182 TATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTA 3241
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 TATGCCAATCAGCAAGGCTGGACGCTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTA 720

Qy       3242 CCGATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGC 3301
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CCGATTACCCTGAGCGGCAACATCGATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGC 780

Qy       3302 AGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTG 3361
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGCGGCCTGCGCGCCGGGCTGTTTCATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTG 840

Qy       3362 GCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACAT 3421
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 GCGCTAGCGGAGTTGGTAGCAGCGCTGTTGCCCACCGATGCCCTCACGCTCAGCCAACAT 900

Qy       3422 ATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGC 3481
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ATCGAACGCTTTGCCCGTCAGCAGTGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGC 960

Qy       3482 ATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGG 3541
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATGCTGTTTTTGGCCGGTAAGCCGCAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGG 1020

Qy       3542 CTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACG 3601
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 CTCGATGCCGGGTTAATTAGCCGCTTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACG 1080

Qy       3602 CGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAAT 3661
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CGGATTCTGTGCGGCAAGCCGCCGGTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAAT 1140

Qy       3662 TCTGTCGAACCAGGGAAGAAAAAATGA 3688
              |||||||||||||||||||||||||||
Db       1141 TCTGTCGAACCAGGGAAGAAAAAATGA 1167

RESULT 14
US-10-804-677-7
; Sequence 7, Application US/10804677
; Publication No. US20040224383A1
; GENERAL INFORMATION:
;  APPLICANT: E.I. duPont de Nemours and Company, Inc.
;  APPLICANT:  Cheng, Qiong
;  APPLICANT:  Tao, Luan
;  APPLICANT:  Sedkova, Natalia
;  TITLE OF INVENTION: GENES ENCODING CAROTENOID COMPOUNDS
;  FILE REFERENCE: CL2346 US NA
;  CURRENT APPLICATION NUMBER: US/10/804,677
;  CURRENT FILING DATE:  2004-03-19
;  PRIOR APPLICATION NUMBER: US 60/468,596
;  PRIOR FILING DATE: 2003-05-07
;  PRIOR APPLICATION NUMBER: US 60/527,083
;  PRIOR FILING DATE: 2003-12-03
;  NUMBER OF SEQ ID NOS: 20
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7
;   LENGTH: 1482
;   TYPE: DNA
;   ORGANISM: Pectobacterium cypripedii DC416
US-10-804-677-7

  Query Match             15.5%;  Score 1085.2;  DB 18;  Length 1482;
  Best Local Similarity   83.3%;  Pred. No. 0;
  Matches 1234;  Conservative    0;  Mismatches  248;  Indels    0;  Gaps    0;

Qy       3685 ATGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTG 3744
              ||||||||||| ||||||||||||||||||||||| || |||||||||||||||||| |||
Db          1 ATGAAACGCACCTATGTGATTGGCGCAGGCTTCGGTGGCCTGGCGCTGGCGATTCGTCTG 60

Qy       3745 CAAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCC 3804
              |||||||| ||| | ||   |||  | || || |||||||| || || ||||| ||||||
Db         61 CAAGCGGCCGGCGTGCCGGTCACGCTGCTGGAACAGCGCGATAAGCCTGGCGGGCGCGCC 120

Qy       3805 TATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCC 3864
              ||||||| | ||||    || |||||||| |||||||| || |||||||| |||||||||
Db        121 TATGTGTATCAGGATCAGGGTTTTACCTTTGATGCCGGTCCGACGGTGATTACCGATCCC 180

Qy       3865 AGCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTG 3924
              ||||| ||||| | | |||| ||||||||||| ||    ||||| ||||| ||||| |||
Db        181 AGCGCTATCGAGGCGCTGTTTACGCTGGCAGGCAAGCAACTCAGTGATTATGTCGACCTG 240

Qy       3925 ATGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGAC 3984
              |||||||| ||||| || |||||||||||||||||||| ||||  ||||| || ||||||
Db        241 ATGCCGGTGACGCCATTTTATCGCCTGTGCTGGGAAGACGGCAGGCAGCTGGACTACGAC 300

Qy       3985 AATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGC 4044
              || |||||| ||| |||||||||||||||| ||||| || |||| || |||||| |  || 
Db        301 AACAATCAGGCGCAGCTGGAGCAGCAGATTGCCACTTTTAATCCCCAGGATGTCGCCGGT 360

Qy       4045 TATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGC 4104
              || || || ||||| |||||||||||  || || ||| | ||||| ||||||||||| |||
Db        361 TACCGCCAGTTTCTGGCCTATTCACAGGATGTGTTTCGTGAGGGCTATCTGAAACTGGGC 420

Qy       4105 ACGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTG 4164
              || || || ||||||||  | || |||||||||||| | || ||| ||| |||| || |||
Db        421 ACCGTACCTTTTCTGCATTTCCGCGACATGCTGCGTGCCGGGCCACAGCTGGGTCGGCTG 480

Qy       4165 CAAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGT 4224
              || || |||||||||| ||||||||||||||||||||||||||||||| |||||||||||||| 
Db        481 CAGGCCTGGCGCAGTGTCTACAGCATGGTGGCGAAATTTATTCATGACGATCATCTGCGC 540

Qy       4225 CAGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATC 4284
              ||||| ||||||||| |||||| |||||||| |||||||||||||||||||||||| ||||||
Db        541 CAGGCTTTTTCCTTTCACTCGTTGCTGGTCGGCGGTAATCCTTTTGCAACGTCTTCGATC 600

Qy       4285 TATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACC 4344
              |||||||||||||| || |||||||| ||||||||||||||||||||||||||||| |||
Db        601 TATACCTTAATTCACGCACTGGAGCGCGAATGGGGCGTGTGGTTTCCGCGCGGCGGTACC 660

Qy       4345 GGCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTG 4404
              || ||||||||  | ||||||||||| |||||    || ||||||||| ||| | || 
Db        661 GGTGCGCTGGTTGATGGCATGGCGCGGCTGTTTCGCGATTTGGGCGGTGAACTGCTGCTC 720

Qy       4405 AATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAG 4464
              || ||||||||| |||||||||||| |||   || || |||| ||||| || ||| |  ||
Db        721 AACGCCGAAGTCAGCCAGCTGGAGACCGAGGGTAACCGCATCAGCGGTGTCCAGCTGAAG 780

Qy       4465 GGCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTAC 4524
              |  || || || || |  |||||||| || || || ||||| ||||||||||||||||||
Db        781 GATGGGCGCCGTTTTGCCGCCGCCGCCGTTGCGTCAAATGCTGACGTGGTGCATACCTAC 840

Qy       4525 GACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAG 4584
              ||    ||| |  |||| |||||   |||    ||||| ||| || |||||||||| |||
Db        841 GATCGCCTGTTAAGCCAGCATCCTGCGGCGCGTAAACGCGCGGCAACGCTGAAGCGCAAG 900

Qy       4585 CGCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTC 4644
              || |||||||||||||||||||||||||||||||||| || |||||  | ||    ||||| 
Db        901 CGGATGAGCAACTCGCTGTTTGTACTCTATTTTGGTCTTAATCATGCCCACCCGCAGCTG 960
```

```
Qy        4645 GCGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAAC 4704
               ||||||||||| || || ||||| ||||| |||||||| |||||||||||||| |||||
Db         961 GCGCACCACACGGTGTGCTTTGGTCCGCGCTATCGTGAATTGATCGATGAGATCTTCAAT 1020

Qy        4705 AGCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCG 4764
               |||||||||||||| || ||||| || || || ||||| |||||||||  ||||||||||
Db        1021 AGCAGCCAGCTGGCGGAAGATTTCTCGCTGTATCTGCATGCGCCCTGCTCCAGCGATCCG 1080

Qy        4765 TCGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGC 4824
               |||||||||||| | ||||||||||| ||||| ||| | ||||||||||||||||||||
Db        1081 TCGCTGGCACCGGCGGGCTGCGGCAGTTTTTACGTGCTGGCGCCGGTGCCGCATCTCGGT 1140

Qy        4825 ACCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTAT 4884
               ||||| |  || ||||||||||||||||| ||||||||||||||||| ||||||||||||
Db        1141 ACCGCCGCAATTGACTGGCAACAGGAAGGGCCGCGCTTGCGCGATCGCATTTTTGCTTAT 1200

Qy        4885 CTGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACG 4944
               |||||| ||||||| ||||||||  | || ||||| |||||||||||| | |||||||||
Db        1201 CTGGAGGAGCACTATATGCCGGGTCTGCGACAGCAGTTAGTGACACACCGTATGTTTACG 1260

Qy        4945 CCGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCG 5004
               |||||||||||||||||||||||||| || ||||| ||||| ||||||||||| || ||
Db        1261 CCGTTTGATTTTCGCGACACGCTGCACGCGCATCAGGGCTCAGCGTTTTCGCTCGAACCC 1320

Qy        5005 ATTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTC 5064
               |||||||||||||||||||||||||| |||||||||||||||||||| || |  || ||
Db        1321 ATTTTGACGCAAAGCGCCTGGTTCCGGCCGCATAACCGCGATGCCGACATTACTAACCTT 1380

Qy        5065 TATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCC 5124
               ||||||||||| || || |||||||| || || || ||||| ||||||||||| || ||
Db        1381 TATCTGGTGGGGGCTGGCACGCATCCCGGTGCCGGTGTGCCAGGCGTGATCGGCTCCGCG 1440

Qy        5125 AAGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGA 5166
               || || ||||||  ||||||| |||||||||   ||| ||||
Db        1441 AAAGCGACCGCCCAGCTGATGGTGGAGGATCTGACCGGATGA 1482


RESULT 15
US-09-920-923-1/c
; Sequence 1, Application US/09920923
; Publication No. US20030022273A1
; GENERAL INFORMATION:
;  APPLICANT: Pasamontes, Luis
;  APPLICANT:  Tsygankov, Yuri
;  TITLE OF INVENTION: Improved Fermentative Carotenoid Production
;  FILE REFERENCE: Improved Fermentive Carotenoid
;  CURRENT APPLICATION NUMBER: US/09/920,923
;  CURRENT FILING DATE:  2001-08-02
;  PRIOR APPLICATION NUMBER: 08/980,832
;  PRIOR FILING DATE: 1997-12-01
;  NUMBER OF SEQ ID NOS: 66
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 8625
;   TYPE: DNA
;   ORGANISM: Flavobacterium sp. R1534
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (8348)..(8349)
;   NAME/KEY: unsure
;   LOCATION: (8539)..(8540)
;   NAME/KEY: unsure
;   LOCATION: (8581)
;   NAME/KEY: unsure
;   LOCATION: (8590)
;   NAME/KEY: unsure
;   LOCATION: (8592)
;   NAME/KEY: unsure
;   LOCATION: (8602)..(8604)
US-09-920-923-1

  Query Match              14.8%;  Score 1035.4;  DB 10;  Length 8625;
```

```
  Best Local Similarity   56.5%;  Pred. No. 3.8e-309;
  Matches 2037;  Conservative    0;  Mismatches 1536;  Indels   30;  Gaps    5;

Qy        2486 GTGCCAGCAGCAAGTCGTGCTGGCGGAGGCGACCTGATGCGCACGCAATACGATGTGATT 2545
               | ||  || ||     |  |  |  | ||||    |     || |    | ||| || |
Db        6984 GGGCGCGCTGCTGAAGGACCGCGAAGGGGCGGATCGCAATACATGAGCCATGATCTGCTG 6925

Qy        2546 TTGGTCGGTGCTGGACTGGCGAATGGCTTGATTGCGCTGCGTCTGCGTCAATTGCAGCCA 2605
                | |  || || || ||| |   ||   |||| |||||     | ||  |       ||
Db        6924 ATCGCGGGCGCGGGGCTGTCCGGTGCGCTGATCGCGCTTGCCGTTCGCGACCGCAGACCG 6865

Qy        2606 CAACTGAAATGCCTGTTGCTGGAGAGCGATGCGCATCCGGCAGGCAATCATACCTGGTCG 2665
                |   |     | || |||| ||        |    ||  | | | | || ||||||||
Db        6864 GATGCGCGCATCGTGATGCTCGACGCGCGGTCCGGCCCCTCGGACCAGCACACCTGGTCC 6805

Qy        2666 TTTCATCACAGCGATCTCAGCGCCGAACAACTTCGCTGGCTGCAACCGCTGATTACCGTG 2725
               |  ||  |||  |||||     |||||   ||     | |||   ||  |   |  ||
Db        6804 TGCCACGACACGGATCTTTCGCCCGAATGGCTGGCGCGCCTGTCGCCCATTCGTCGCGGC 6745

Qy        2726 CGTTGGTCAGGTTATCAGGTGCGTTTTCCTGCGCTGCGCCGCAATCTGGACGGGGATTAT 2785
                  ||| | | | |  ||||    ||||| |  |     |||   |||      |  |||
Db        6744 GAATGGACGGATCAGGAGGTCGCGTTTCCCGACCATTCGCGCCGCCTGACGACAGGCTAT 6685

Qy        2786 TGTTCCATCGCATCAGGCGATTTTGCCCGCCATCTTTACGCGGCGATGGGTGACGATCTG 2845
                | || ||||   | ||||   |   | | |  ||  |         ||||| |||||||
Db        6684 GGCTCGATCGAGGCGGGCGCGCTGATCGGGCTGCTGCA---------GGGTGTCGATCTG 6634

Qy        2846 TGGACAAACACAGCCGTACAACAGGTAAAACCCACGCAGGTGACGCTGGCGGATGGCCGT 2905
                ||   || ||    ||      | |  |    ||    | ||||||| |||| |||
Db        6633 CGGTGGAATACGCATGTCGCGACGCTGGACGATACCGGCGCGACGCTGACGGACGGCTCG 6574

Qy        2906 GAACTTGCTGCGCAAGTGGTGATTGATGGTCGCGGCCTGCAGCCGACGCCACATCTGCAG 2965
                   | |  ||       ||||| || |  || ||         ||| || || |||
Db        6573 CGGATCGAGGCTGCCTGCGTGATCGACGCCCGTGGTGCCGTCGAGACCCCGCACCTGACC 6514

Qy        2966 CTGGGTTATCAGGTGTTTCTTGGACAAGAGTGGCAGCTGGCGCAGCCGCACGGCCTGCAG 3025
                ||||||  |||   ||  | ||    |||    ||   |     || || ||| |  ||
Db        6513 GTGGGTTTCCAGAAATTCGTGGGCGTCGAGATCGAGACCGACGCCCCCCATGGCGTCGAG 6454

Qy        3026 CAGCCGATCCTGATGGATGCCACCGTCGATCAGCAAGCGGGTTATCGTTTTGTCTACACG 3085
               |  |||||  | ||||| || |||||    |||   |  || || || ||   ||||   |
Db        6453 CGCCCGATGATCATGGACGCGACCGTTCCGCAGATGGACGGGTACCGCTTCATCTATCTG 6394

Qy        3086 CTGCCGCTCAGCGCCGATCGGCTATTGATTGAAGATACCCATTACGTTAACCAGCCCGCG 3145
               |||||  ||||  ||   ||  |  |||| || ||||| |  |||    ||     ||
Db        6393 CTGCCCTTCAGTCCCACCCGCATCCTGATCGAGGATACGCGCTACAGCGACGGCGGCGAT 6334

Qy        3146 CTGGCGGAGAACACCGCTCGTCAGCACATCGCCGACTATGCCAATCAGCAAGGCTGGACG 3205
               ||||  ||   | |       |||         |||||||||     ||  ||||||||
Db        6333 CTGGACGATGGCGCGCTGGCGCAGGCGTCGCTGGACTATGCCGCCAGGCGGGGCTGGACC 6274

Qy        3206 CTGAGTACGCTGCTGCGTGAAGAGCACGGCATATTACCGATTACCCTGAGCGGCAACATC 3265
                 |     | ||| ||| |||  |  |  | |                   | |  |||
Db        6273 GGGCAGGAGATGCGGCGCGAAAGGGGCATCCTGCCCATCGCGCTGGCCCATGACGCCATA 6214

Qy        3266 GATCGATTCTGGCAACAGCAGCGCGGCCAAGCGTGCAGCGGCCTGCGCGCCGGGCTGTTT 3325
               |     |   |  | ||   ||  ||    | |      || ||| | || ||||||||
Db        6213 GGCTTCTGGCGCGACCACGCGCAGGGGGCGGTGCCGGTTGGGCTGGGGGCAGGGCTGTTC 6154

Qy        3326 CATGCCACCACCGGTTACTCCTTGCCGTCCGCCGTGGCGCTAGCGGAGTTGGTAGCAGCG 3385
               ||  ||  |||||| || ||  |||| |  |||| |  | | |||||     | || |||
Db        6153 CACCCCGTCACCGGATATTCGCTGCCCTATGCCGCGCAGGTCGCGGATGCCATCGCGGCG 6094

Qy        3386 CTGTTGCCCACCGATGCCCTCACGCTCAGCCAACATATCGAACGCTTTGCCCGTCAGCAG 3445
               |     |  ||    ||   | | |   ||        ||   |      |  ||    |
Db        6093 CGCGACCTGACGACCGCGTCCGCCCGTCGCGCGGTGCGCGGCTGGGCCATCGATCGCGCG 6034

Qy        3446 TGGCGCGAACAGCGATTTTTCCGTCTGCTAAACCGCATGCTGTTTTTGGCCGGTAAGCCG 3505
                  ||||   | || ||  | || ||||| ||||| ||||||||    | | |   |||
Db        6033 GATCGCG---ACCGCTTCCTGCGGCTGCTGAACCGGATGCTGTTCCGCGGCTGCCCGCCC 5977

Qy        3506 CAGCAGCGCTGGCGCGTGATGCAACGTTTTTACCGGCTCGATGCCGGGTTAATTAGCCGC 3565
                | |  ||||  ||| || |||| || || ||||| ||         | | ||    |||
Db   5976 GACCGTCGCTATCGCCTGCTGCAGCGGTTCTACCGCCTGCCGCAGCCGCTGATCGAGCGC 5917

Qy   3566 TTTTACGCCGGGCAACTGCGCCTGCGCGATAAAACGCGGATTCTGTGCGGCAAGCCGCCG 3625
          || || ||||||||  |||    ||  |||       || ||  |  |   |||    |||||
Db   5916 TTCTATGCCGGGCGCCTGACATTGGCCGACCGGCTTCGCATCGTCACCGGACGCCCGCCC 5857

Qy   3626 GTGCCCATCGGTGAAGCGCTGCGCGCGCTGTTGAATTCTGTCGAACCAGGGAAGAAAAAA 3685
           | ||  |     | ||  |||||   |||    |      |   |    | ||| |  |
Db   5856 ATTCCGCTGTCGCAGGCCGTGCGCTGCCTGCCCGAACGCCCCCTGCTGCAGGAGAGAGCA 5797

Qy   3686 TGAAACGCACTTATGTGATTGGCGCAGGCTTTGGCGGCCTGGCGCTGGCGATTCGCCTGC 3745
          |||    | |    || || |||||||| || ||||| || ||||| || || |||||||
Db   5796 TGAGTTCCGCCATCGTCATCGGCGCAGGTTTCGGCGGGCTTGCGCTTGCCATCCGCCTGC 5737

Qy   3746 AAGCGGCGGGCATACCAACCACCTTACTCGAGCAGCGCGACAAACCGGGCGGACGCGCCT 3805
          || |||| ||||||  | |||||| |  |||||   |||||||| || ||||| |||||||
Db   5736 AATCGGCCGGCATCGCGACCACCATCGTCGAGGCCCGCGACAAGCCCGGCGGCCGCGCCT 5677

Qy   3806 ATGTGTTTGAGGACAGTGGCTTTACCTTCGATGCCGGACCCACGGTGATCACCGATCCCA 3865
          |||| |   | ||    |||     ||||||||| || || |||||  | ||||| |||
Db   5676 ATGTCTGGAACGATCAGGGCCACGTCTTCGATGCAGGCCCGACGGTCGTGACCGACCCCG 5617

Qy   3866 GCGCCATCGAAGAGTTGTTCACGCTGGCAGGAAAATCGCTCAGCGATTACGTCGAGCTGA 3925
           |  | |   |||| |||    | ||    ||  || || |      | ||||   ||||
Db   5616 ACAGCCTGCGAGAGCTGTGGGCCCTCAGCGGCCAACCGATGGAGCGTGACGTGACGCTGC 5557

Qy   3926 TGCCGGTAACGCCCTTCTATCGCCTGTGCTGGGAAGATGGCAAACAGCTTGATTACGACA 3985
          |||||||  ||||||||||| || |||   ||||  || |||       | || ||||  |
Db   5556 TGCCGGTCTCGCCCTTCTACCGGCTGACATGGGCGGACGGCCGCAGCTTCGAATACGTGA 5497

Qy   3986 ATAATCAGCCGCTGCTGGAGCAGCAGATCGCCACGTTCAATCCGCAAGATGTAGAAGGCT 4045
          |  |  |      ||||   |  ||| ||||| | |||||||    ||||| || ||||
Db   5496 ACGACGACGACGAGCTGATCCGCCAGGTCGCCTCCTTCAATCCCGCCGATGTCGATGGCT 5437

Qy   4046 ATCGTCAATTTCTTGCCTATTCACGTGAAGTATTTAGAGAGGGTTATCTGAAACTCGGCA 4105
          |||| |  || |  |  ||  |    || || | | | | |||| |||||||| || || |
Db   5436 ATCGCCGCTTCCACGATTACGCCGAGGAGGTCTATCGCGAGGGGTATCTGAAGCTGGGGA 5377

Qy   4106 CGGTGCCGTTTCTGCAGGTGCGTGACATGCTGCGCGTCGCGCCGCAGTTGGGACGTCTGC 4165
          |   ||| || ||| || || |  | ||||||  || |||||||  | ||   || ||||
Db   5376 CCACGCCCTTCCTGAAGCTGGGCCAGATGCTGAACGCCGCGCCGGCGCTGATGCGCCTGC 5317

Qy   4166 AAGCATGGCGCAGCGTCTACAGCATGGTGGCGAAATTTATTCAGGACGATCATCTGCGTC 4225
          | ||||  |||   ||| ||||||||||||||   || || ||||||   |||||||| |
Db   5316 AGGCATACCGCTCGGTCCACAGCATGGTGGCGCGCTTCATCCAGGACCCGCATCTGCGGC 5257

Qy   4226 AGGCGTTTTCCTTCCACTCATTGCTGGTGGGCGGTAATCCTTTTGCAACGTCATCGATCT 4285
          |||| || || |||||| |  ||||||| ||||| || || ||| | ||    |||||||
Db   5256 AGGCCTTCTCGTTCCACACGCTGCTGGTCGGCGGGAACCCGTTTTCGACCAGCTCGATCT 5197

Qy   4286 ATACCTTAATTCATGCGCTGGAGCGTGAATGGGGCGTGTGGTTTCCGCGCGGCGGCACCG 4345
          || |  | || ||||||||||| ||     | ||||| ||||||  |    |||||||||
Db   5196 ATGCGCTGATCCATGCGCTGGAACGGCGCGGCGGCGTCTGGTTCGCCAAGGGCGGCACCA 5137

Qy   4346 GCGCGCTGGTGCAGGGCATGGCGCGACTGTTCGAGGACTTGGGCGGCGAGCTGTTACTGA 4405
           |  ||||||   |||||||| |     |||||||||    | ||||||  |||| | ||||
Db   5136 ACCAGCTGGTCGCGGGCATGGTCGCCCTGTTCGAGCGTCTTGGCGGCACGCTGCTGCTGA 5077

Qy   4406 ATGCCGAAGTGAGCCAGCTGGAAACCAGCGGCAATCGCATTAGCGGCGTTCAGTTAGAGG 4465
          |||||   || |  | | | || |||   ||| |||||   |  |||||   | |   ||
Db   5076 ATGCCCGCGTCACGCGGATCGACACCGAGGGCGATCGCGCCACGGGCGTCACGCTGCTGG 5017

Qy   4466 GCGGACGACGCTTCGATGCCGCCGCTGTGGCCTCCAATGCCGACGTGGTGCATACCTACG 4525
           ||| || |  ||    || |   | ||||||  |||  | | ||||||| |||| | |||
Db   5016 ACGGGCGGCAGTTGCGCGCGGATACGGTGGCCAGCAACGGCGACGTGATGCACAGCTATC 4957

Qy   4526 ACAAACTGCTTCGCCACCATCCGCTGGCAATGAAACGTGCGACATCGCTGAAGCGTAAGC 4585
           | | |||||  ||||    |  |  |     |     ||  |    ||||| ||  |||
Db   4956 GCGACCTGCTGGGCCATACCCGCCGCGGGCGCACCAAGGCCGCGATCCTGAACCGGCAGC 4897

Qy   4586 GCATGAGCAACTCGCTGTTTGTACTCTATTTTGGCCTGAATCAGCCGCATGAACAGCTCG 4645
          ||  |   |  |||||||| || ||  ||||| ||||||    |||  |  ||  | || |
Db   4896 GCTGGTCGATGTCGCTGTTCGTGCTGCATTTCGGCCTGTCCAAGCGCCCCGAGAACCTGG 4837

Qy   4646 CGCACCACACCGTCTGTTTTGGCCCGCGTTATCGTGAGTTGATCGATGAGATTTTCAACA 4705
          | ||||||| ||||   || |||||||| ||    | | || |  | ||||| |||||||
Db   4836 CCCACCACAGCGTCATCTTCGGCCCGCGCTACAAGGGGCTGGTGAACGAGATCTTCAACG 4777

Qy   4706 GCAGCCAGCTGGCAGACGATTTTTCACTTTACCTGCACGCGCCCTGCAGCAGCGATCCGT 4765
          |    |  ||| | |||||||| ||  | |  | |||||  |||||||   | ||||||
Db   4776 GGCCACGCCTGCCGGACGATTTCTCGATGTATCTGCATTCGCCCTGCGTGACCGATCCCA 4717

Qy   4766 CGCTGGCACCGCCCGGCTGCGGCAGCTTTTATGTGTTAGCGCCGGTGCCGCATCTCGGCA 4825
            ||||| ||    ||  |   ||    ||| ||  | ||||| || ||||||||| |||
Db   4716 GCCTGGCCCCCGAGGGGATGTCCACGCATTACGTCCTTGCGCCCGTTCCGCATCTGGGCC 4657

Qy   4826 CCGCTGACATCGACTGGCAACAGGAAGGACCGCGCTTGCGCGATCGAATTTTTGCTTATC 4885
           ||| ||  |||| ||| ||   || |  ||| |||    ||| || || || |   | |
Db   4656 GCGCCGATGTCGATTGGGAAGCCGAGGCCCCGGGCTATGCCGAGCGCATCTTCGAGGAAC 4597

Qy   4886 TGGAGCAGCACTACATGCCGGGATTACGTCAGCAATTAGTGACACACAGAATGTTTACGC 4945
          ||||||  | |  ||| || |   | ||  ||||  |         | | || || |  |
Db   4596 TGGAGCGCCGCGCCATCCCCGACCTGCGCAAGCACCTGACCGTCAGCCGCATCTTCAGCC 4537

Qy   4946 CGTTTGATTTTCGCGACACGCTGCATGCCCATCACGGCTCGGCGTTTTCGCTGGAGCCGA 5005
          |    |||||  ||  |   |||   ||||||||||||   || || ||| | |||||||
Db   4536 CCGCCGATTTCAGCACCGAACTGTCGGCCCATCACGGCAGCGCCTTCTCGGTCGAGCCGA 4477

Qy   5006 TTTTGACGCAAAGCGCCTGGTTCCGCCCGCATAACCGCGATGCCGATATCAGCAATCTCT 5065
          |  |||||||||  |||||||||||||||||||||||||||   ||  |||   ||  |||
Db   4476 TCCTGACGCAATCCGCCTGGTTCCGCCCGCATAACCGCGACCGCGCGATCCCGAACTTCT 4417

Qy   5066 ATCTGGTGGGTGCCGGTACGCATCCAGGCGCGGGCGTGCCCGGCGTGATCGGTTCGGCCA 5125
          |  | ||||| || || ||||||| || ||||||| | || || ||  | ||    ||||
Db   4416 ACATCGTGGGGGCGGGCACGCATCCGGGTGCGGGCATCCCGGGTGTCGTTGGCAGCGCCA 4357

Qy   5126 AGGCCACCGCCAGGCTGATGCTGGAGGATCGCGCCGAATGAATCGACAGCCTTTACTTGA 5185
          ||||||| ||   | | ||||||  ||| |  ||||    |      | |
Db   4356 AGGCCACGGCGCAGGTCATGCTGTCGGACCTGGCCGTCGCATGACCGATCTGACGGCGAC 4297

Qy   5186 GCAAGTAACGCAAACCATGGCGGTGGGCTCGAAGAGTTTCGCCACCGCCGCCAAGCTGTT 5245
              | | ||    |||| |||  ||| ||| | || |||||    || |||||||| |
Db   4296 TTCCGAAGCG---GCCATCGCGCAGGGTTCGCAAAGCTTCGCGCAGGCGGCCAAGCTGAT 4240

Qy   5246 TGATGCACCGACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGA 5305
               |    |  |||      ||| | |||| ||||| |||||  | |    |||||
Db   4239 GCCGCCCGGCATCCGCGAGGATACGGTCATGCTCTATGCCTGGTGCAGGCATGCGGATGA 4180

Qy   5306 TGTGATTGATGGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCA 5365
           ||||| || |||||   | ||||    | |  |  |     || | ||  ||| | ||
Db   4179 CGTGATCGACGGGCAGGTGATGGGTTCTGCCCCCGAG------GCGGGCGGCGACCCACA 4126

Qy   5366 GGCACGTATGCAGCATCTGCAAATTGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGA 5425
          ||| ||  ||  |   ||||     || || |    |||      ||  |    |  |
Db   4125 GGCGCGGCTGGGGGCGCTGCGCGCCGACACGCTGGCCGCGCTGCACGAGGACGGCCCGAT 4066

Qy   5426 TGAACCGGCGTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACT 5485
              ||| | ||   |||| | |   | || ||       ||  |  | | || | |
Db   4065 GTCGCCGCCCTTCGCGGCGCTGCGCCAGGTCGCCCGGCGGCATGATTTCCCGGACCTTTG 4006

Qy   5486 GGCGTTTGATCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTT 5545
          | || | || |   | || || ||||| ||||||||       |     | ||||   ||| |
Db   4005 GCCGATGGACCTGATCGAGGGTTTCGCGATGGATGTCGCGGATCGCGAATACCGCAGCCT 3946

Qy   5546 CGATGACACGCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCG 5605
           ||||||  ||||   || | ||| |||||||||||| || || ||  |||||||||||||
Db   3945 GGATGACGTGCTGGAATATTCCTACCACGTCGCGGGGGTCGTGGGCGTGATGATGGCGCG 3886

Qy   5606 CGTAATGGGCGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTT 5665
           || ||||||||||  ||||| |||||||| |||| |||||||||| | || || |||||
Db   3885 GGTGATGGGCGTGCAGGACGATGCGGTGCTGGATCGCGCCTGCGATCTGGGCCTTGCGTT 3826

Qy   5666 CCAGCTCACTAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCT 5725
          |||||| || ||||| || ||||| |  | || ||||||||  |  || |||||||||||
Db   3825 CCAGCTGACGAACATCGCTCGCGACGTGATCGACGATGCCGCCATCGGGCGCTGCTATCT 3766
```

```
Qy        5726 GCCGCAATCCTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACA 5785
               |||      |||||| |   |||||||          |  ||| |    |  |||   |
Db        3765 GCCTGCCGACTGGCTGGCCGAGGCGGG---------GGCGACGGTTGAGGGTCCGGTGCC 3715

Qy        5786 TCGTGCAGCGCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCA 5845
               |   |  ||||||   ||  |       ||  |  | |  | ||| || |||||||||
Db        3714 TTCGGACGCGCTCTATTCCGTCATCATCCGCCTGCTTGACGCGGCCGAGCCCTATTATGC 3655

Qy        5846 CTCGGCGCGATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGC 5905
               |||||||||    ||  | |||  | | |||| |||||  |||||| | |||||  | ||
Db        3654 CTCGGCGCGGCAGGGGCTTCCGCATCTGCCGCCGCGCTGCGCGTGGTCGATCGCCGCCGC 3595

Qy        5906 TCGCGGCGTTTATCGCGAAATTGGCGTCAAAGTTCAGCACGCCGGTGTGCACGCCTGGGA 5965
                |   |  | ||||||| ||| ||        | | ||| |  ||     | ||||
Db        3594 GCTGCGTATCTATCGCGCAATCGGGACGCGCATCCGGCAGGGTGGCCCCGAGGCCTATCG 3535

Qy        5966 TTCACGGCAGCGCACCAGTAAAGGTGAAAAACTGGCGCTGCTGGTGAAAGGGGCAGGTTT 6025
                   |||    ||||    || | ||  ||  | | ||| |||| |   || |
Db        3534 CCAGCGGATCAGCACGTCGAAGGCTGCCAAGATCGGGCTTCTGGCGCGCGGAGGCTTGGA 3475

Qy        6026 GGCGATCACTTCGCGTGTGTCTCGTCCTGAACCGCGTCCGGCTGGTCTGTGGCAGCGTCC 6085
                |||  | | |||||  ||    |    |||    | |  |  || ||||||   || ||
Db        3474 CGCGGCCGCATCGCGCCTGCGCGGCGGCGAAATCAGCCGCGACGGCCTGTGGACCCGACC 3415

Qy        6086 TCG 6088
                ||
Db        3414 GCG 3412

Search completed: January 26, 2005, 02:16:28
Job time : 3368 secs
```

```
                         GenCore version 5.1.6
                Copyright (c) 1993 - 2005  Compugen Ltd.


OM nucleic - nucleic search, using sw model

Run on:         January 24, 2005, 22:53:11 ; Search time 19359 Seconds
                                          (without alignments)
                                          13174.321 Million cell updates/sec

Title:          US-10-808-979-18
Perfect score:  6999
Sequence:       1 gcgatggcaatggaaaatgt..........gacgttttagacatgcaata 6999

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters:      65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                1:  gb_est1:*
                2:  gb_est2:*
                3:  gb_htc:*
                4:  gb_est3:*
                5:  gb_est4:*
                6:  gb_est5:*
                7:  gb_est6:*
                8:  gb_gss1:*
                9:  gb_gss2:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                              SUMMARIES
                    %
Result              Query
  No.     Score     Match Length DB  ID                    Description
 ------------------------------------------------------------------------
c   1     231.6     3.3     659  8  AQ989968              AQ989968 Rfc00652
c   2     131.4     1.9     621  8  AQ990546              AQ990546 Rfc01339
c   3      77.6     1.1     315  6  CB041878              CB041878 4006842 B
c   4      77.6     1.1     315  6  CB042040              CB042040 4007005 B
c   5      77.6     1.1     349  6  CB041738              CB041738 4006549 B
    6      77.6     1.1     359  6  CB041616              CB041616 4006367 B
    7      77.4     1.1     314  6  CB061284              CB061284 4011834 B
c   8      75.8     1.1     300  6  CB061218              CB061218 4011666 B
c   9      75.8     1.1     317  6  CB059804              CB059804 4009880 B
c  10      74.4     1.1     354  6  CB041601              CB041601 4006361 B
c  11 .    74.2     1.1     302  6  CB062064              CB062064 4012128 B
c  12      73.6     1.1     475  8  BH439236              BH439236 BOHAC18TF
   13      73.6     1.1     754  8  BZ434062              BZ434062 BONFQ34TF
   14        72     1.0     425  2  BE323612              BE323612 NF006A09P
   15        72     1.0     709  8  BH502634              BH502634 BOHJK46TR
   16        72     1.0     736  6  CA916997              CA916997 EST641144
c  17      71.8     1.0     775  8  BZ550668              BZ550668 pacs1-60_
   18      71.6     1.0     809  7  CF517921              CF517921 CAP0005_I
   19      71.6     1.0     833  7  CF513004              CF513004 CAbud0005
   20      71.2     1.0     632  7  CN144185              CN144185 WOUND1_20
   21      71.2     1.0     725  7  CN144663              CN144663 WOUND1_23
   22      70.8     1.0    1218  8  BZ575772              BZ575772 msh2_462.
   23      70.4     1.0     682  5  BQ856404              BQ856404 QGB3g04.y
   24        70     1.0     453  1  AI442185              AI442185 sa58c11.y
   25      69.6     1.0     698  7  CN148357              CN148357 WOUND1_55
c  26      69.6     1.0     798  8  BZ563162              BZ563162 pacs2-164
   27      69.2     1.0     761  9  CC648441              CC648441 OGUJM19TH
   28      69.2     1.0     777  9  CC722834              CC722834 OGUDU11TH
c  29      69.2     1.0     807  9  CC648445              CC648445 OGUJM19TV
   30      69.2     1.0     929  9  CG441377              CG441377 OGVHI32TV
   31      69.2     1.0    1090  8  BZ553098              BZ553098 pacs1-60_
   32      68.6     1.0     666  4  BI717621              BI717621 1031021A0
   33      68.6     1.0     675  4  BI723806              BI723806 1031067H0
   34      68.6     1.0     701  2  BE238154              BE238154 894040E07
c  35      68.4     1.0     601  6  CB042842              CB042842 4007706 B
   36      68.4     1.0     744  9  CC602991              CC602991 OGUJB14TH
   37      68.4     1.0     785  9  CC661270              CC661270 OGMBA03TH
c  38      68.2     1.0     771  8  BZ442823              BZ442823 BONHG57TR
   39      68.2     1.0     790  7  CF443389              CF443389 EST679734
   40      67.6     1.0     824  6  CB066639              CB066639 EST645262
   41      67.4     1.0     596  8  AQ963747              AQ963747 LERGP06TF
   42        67     1.0     668  4  BI726915              BI726915 1031088E0
   43      66.2     0.9     947  7  CK295442              CK295442 EST758156
c  44      66.2     0.9     975  7  CK295932              CK295932 EST758646
c  45        66     0.9     809  7  CK285594              CK285594 EST748316


                              ALIGNMENTS


RESULT 1
AQ989968/c
LOCUS       AQ989968                 659 bp    DNA     linear   GSS 14-AUG-2000
DEFINITION  Rfc00652 Photorhabdus luminescens strain W14 M13 library
            Photorhabdus luminescens genomic clone PLG00652, genomic survey
            sequence.
ACCESSION   AQ989968
VERSION     AQ989968.1  GI:9648562
KEYWORDS    GSS.
SOURCE      Photorhabdus luminescens
  ORGANISM  Photorhabdus luminescens
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1  (bases 1 to 659)
  AUTHORS   ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
            Daborn,P.J., Bowen,D. and Blattner,F.R.
  TITLE     A genomic sample sequence of the entomopathogenic bacterium
            Photorhabdus luminescens W14: potential implications for virulence
  JOURNAL   Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
  MEDLINE   20378633
   PUBMED   10919786
COMMENT     Contact: ffrench-Constant RH
            Department of Biology and Biochemistry
            University of Bath
            South Building, Bath BA2 7AY, UK
            Tel: (44) 1225 826621
            Fax: (44) 1225 826779
            Email: bssrfc@bath.ac.uk
            This is one of 2,122 random reads from the M13 library. For
            annotation of identified clones (BLASTX, BLASTN and mapping to E.
            coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
            Acids Res.
            Seq primer: M13 Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .659
                     /organism="Photorhabdus luminescens"
                     /mol_type="genomic DNA"
                     /strain="W14"
                     /db_xref="taxon:29488"
                     /clone="PLG00652"
                     /dev_stage="primary phase variant"
                     /clone_lib="Photorhabdus luminescens strain W14 M13
                     library"
                     /note="Genomic DNA from strain W14 was size selected (1-2
                     kb) and then cloned into M13 Janus."
ORIGIN
```

```
  Query Match             3.3%;  Score 231.6;  DB 8;  Length 659;
  Best Local Similarity   62.5%;  Pred. No. 2.5e-51;
  Matches  408;  Conservative    0;  Mismatches  242;  Indels    3;  Gaps    3;

Qy         5256 ACGCGCCGCAGCACGCTGATGCTGTATGCGTGGTGTCGTCACTGCGATGATGTGATTGAT 5315
                || || | |||| |  |||||||||| || ||||   || | ||  ||||| | || ||
Db          653 ACCCGACACAGCNCAATGATGCTGTACGCCTGGTNCNGTTATTG-TATGATATAATAGAC 595

Qy         5316 GGGCAAACGCTGGGCGAAGGCGGCACGCAGCATGCCGTCGAAGACGCGCAGGCACGTATG 5375
                || |||  ||||||  |        |   |        |  ||   |   |  |    |
Db          594 GGCCAAGAGCTGGGAAACAAATAAGCAGTGTTGATAAGCATAGCGCCCGCGAAAAAACTT 535

Qy         5376 CAGCATCTGCAAAT-TGAAACCCGCCGCGCCTACAGCGGCGCGCACATGGATGAACCGGC 5434
                ||    ||||||||     ||  | |  || ||   |||   ||  |||   ||||| ||
Db          534 CAAATGCTGCAAATACCTGACGAGGCAGGCTTATGACGGTTTGCCGATGACCGAACCTGC 475

Qy         5435 GTTTAGGGCGTTTCAGGAAGTGGCGATCATTCACCAGCTGCCGCAACAACTGGCGTTTGA 5494
                 |||   || ||||||  ||  ||  | | | |  |  | || || ||||  || || ||
Db          474 TTTTGCAGCCTTTCAGATAGA-GCTCTGAGTAATGAAATCCCTCAGCAACAAGCCTTCGA 416

Qy         5495 TCATCTGGAAGGCTTCGCTATGGATGCACGCAACGAACATTACGCGAGCTTCGATGACAC 5554
                 ||||||||||| ||||| ||||||| || |   || | ||||  ||  || ||||| ||
Db          415 ACATCTGGAAGGATTCGCAATGGATGTACTCTGTGAGCCTTACAGGACGTTGGATGATAC 356

Qy         5555 GCTGCGTTACTGCTATCACGTCGCGGGCGTGGTCGGTTTGATGATGGCGCGCGTAATGGG 5614
                 ||| | || ||||| || |||||||| ||||| ||  ||||||||||  | || |||||
Db          355 TCTGAGATATTGCTACCATGTCGCGGGAGTGGTTGGGCTGATGATGGCAAGGGTTATGGG 296

Qy         5615 CGTGCGCGACGAAGCGGTGCTCGATCACGCCTGCGATTTAGGACTGGCGTTCCAGCTCAC 5674
                  | |||||       || || ||||  ||||||||| ||||| | || || ||| | ||
Db          295 GATTCGCGAGACGAGTGTACTGGATCGGGCCTGCGATCTAGGAATTGCTTTTCAGTTAAC 236

Qy         5675 TAACATTGCGCGCGACATTGTAGAAGATGCCGAAAATGGTCGCTGCTATCTGCCGCAATC 5734
                ||||||||| || |||||| | || |||||  ||   || || |||||||| |||| |
Db          235 TAACATTGCTCGGGACATTATTGAGGATGCGAAAGCCGGGCGTTGCTATCTTCCGCTAGA 176

Qy         5735 CTGGCTCGATCAGGCGGGATTACGCGCCGATACGCTGACTGCACCGCAACATCGTGCAGC 5794
                 |||||  | || |  ||  ||    | || || || | |    |  || | ||| | ||
Db          175 ATGGCTTTACCAAGAAGGGCTAATGCCAGACACACTTATTTATACCGAAAACCGTCCTGC 116

Qy         5795 GCTCGCCTCACTGGCAGCGCGTTTAGTGGCGGAGGCGGAACCCTATTATCACTCGGCGCG 5854
                ||| ||     ||||| | || ||  | |  || |||||  | ||||||  ||| || |
Db          115 GCTAGCTCGGGTGGCATCCCGGTTGATTGTAGAAGCGGAGTCTTATTATACCTCCGCTCT 56

Qy         5855 ATCCGGTTTACCGGGTTTACCGCTGCGCTCGGCGTGGGCCATCGCTACGGCTC 5907
                  | ||  |    ||||||||  |||| || || |||| ||||||| | ||||
Db           55 GACAGGGCTTGTCGGTTTACCTTTGCGTTCAGCTTGGGTCATCGCTTCTGCTC 3

RESULT 2
AQ990546/c
LOCUS       AQ990546                 621 bp    DNA     linear   GSS 14-AUG-2000
DEFINITION  Rfc01339 Photorhabdus luminescens strain W14 M13 library
            Photorhabdus luminescens genomic clone PLG01339, genomic survey
            sequence.
ACCESSION   AQ990546
VERSION     AQ990546.1  GI:9649140
KEYWORDS    GSS.
SOURCE      Photorhabdus luminescens
  ORGANISM  Photorhabdus luminescens
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Photorhabdus.
REFERENCE   1  (bases 1 to 621)
  AUTHORS   ffrench-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
            Daborn,P.J., Bowen,D. and Blattner,F.R.
  TITLE     A genomic sample sequence of the entomopathogenic bacterium
            Photorhabdus luminescens W14: potential implications for virulence
  JOURNAL   Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
  MEDLINE   20378633
   PUBMED   10919786
COMMENT     Contact: ffrench-Constant RH
            Department of Biology and Biochemistry
            University of Bath
            South Building, Bath BA2 7AY, UK
            Tel: (44) 1225 826621
            Fax: (44) 1225 826779
            Email: bssrfc@bath.ac.uk
            This is one of 2,122 random reads from the M13 library. For
            annotation of identified clones (BLASTX, BLASTN and mapping to E.
            coli K12 genome) please see ffrench-Constant et al. 2000, Nucleic
            Acids Res.
            Seq primer: M13 Forward
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..621
                     /organism="Photorhabdus luminescens"
                     /mol_type="genomic DNA"
                     /strain="W14"
                     /db_xref="taxon:29488"
                     /clone="PLG01339"
                     /dev_stage="primary phase variant"
                     /clone_lib="Photorhabdus luminescens strain W14 M13
                     library"
                     /note="Genomic DNA from strain W14 was size selected (1-2
                     kb) and then cloned into M13 Janus."
ORIGIN

  Query Match             1.9%;  Score 131.4;  DB 8;  Length 621;
  Best Local Similarity   52.0%;  Pred. No. 6e-24;
  Matches  289;  Conservative    0;  Mismatches  264;  Indels    3;  Gaps    1;

Qy          572 AAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATGGATAACGCGCAGA 631
                |||| ||||| || || ||  | ||| |         | ||  |   |   |     ||
Db          599 AAATAGTGCATGCAGCCTCATTAATTTTTANACGATCTTCCNNGCATGGATAATGTGAGT 540

Qy          632 TGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTGGCGATTCTCGCCG 691
                  || |  || |||||     | || ||  | | ||||||||  || ||||| || ||
Db          539 ATCGGCNCGGACGCCCGGNTATTCACCGGCANTATGGTGAAAGTGTAGCGATCCTAGCGT 480

Qy          692 CCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCCGGTTTGCCTGCCA 751
                |  | ||  || | ||||| || ||||   | ||||   | | | |     |  |||
Db          479 CTGTAGCCTTGNTAAGCCGAGCTTTTGGCCTCATTG---TAGAAACAAACCTTTCTGAAT 423

Qy          752 TACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTGCAGGGCTTAGTGC 811
                   ||||||| |  ||  | ||   ||| |||  ||| || ||  |    ||  | || |
Db          422 ACNATAAATCCGCTGCTGTCGCCATACTTTCCCATGCAGTTGGTTTANNCGGTCTTGTCC 363

Qy          812 AAGGGCAATTCCAGGATCTGCACGACGGCACGCAGAGCCGCAGCCCGGAAGCGATCGCCA 871
                ||||||| || |  |||||| || |              | |      | || || || |
Db          362 AAGGGCAGTTTCGCGATCTGAACAATACATTATGTGCTGGTAATTATCANGCCATTGCTA 303

Qy          872 TGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCCACGCTGCAAATGGCGGCGATTG 931
                   ||||| |  | |||||  |  |||| ||   | | ||  | ||| | ||| | ||||
Db          302 CCNCCAACAATTTAAAAACNGGTTTGCTTTTCGACACAACATTTCAATTAGCGNCAATTG 243

Qy          932 CCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGCTTCTTCGCCCAGGATTTGGGCC 991
                | |   | ||  |          || ||   | |  |     |  | ||  || |||| |
Db          242 CTGTGAATGCACCTGATATAACTCGCCAGGCATTACGTCATATATCACAACATCTGGGGC 183

Qy          992 AGGCGTTTCAACTGCTCGACGACCTCGCCGACGGTTGCAAACACACCGGTAAAGATGTGC 1051
                |||||||||||   || ||  || |    || |     ||  | || || |||||| ||
Db          182 AGGCGTTTCAATNACTGGATNACTTGAGTGATGNCCTGAATAATACAGGGAAAGATATGT 123

Qy         1052 ACCAGGATCAGGGCAAATCCACGCTGGTACAGATGCTCGGTGCTGACGGCGCGGAACGTC 1111
                | ||| || ||   || |   |  ||||   |  ||||||          | | |    |
Db          122 ATCAGNATAAGNAAAAGTTANCAATGGTTACGCGGCTCGGAAAATNTANAGTGTATANCC 63

Qy         1112 GCCTGCGCGATCACCT 1127
                 | | |||   ||| |
Db           62 ACTTACGCAGGCACAT 47

RESULT 3
CB041878/c
LOCUS       CB041878                 315 bp    mRNA    linear   EST 15-JAN-2003
```

```
DEFINITION  4006842 BARC-EMBRAPA 315BOV Bos indicus cDNA clone 315BOV_1E16
            Unknown, mRNA sequence.
ACCESSION   CB041878
VERSION     CB041878.1  GI:27761123
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 18
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: E  column: 16
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 315.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="315BOV_1E16"
                     /sex="female"
                     /tissue_type="mammary"
                     /cell_type="epithelium"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 315BOV"
                     /note="Organ: mammary; Vector: pUC118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature, PCR with 16mer AM5."
ORIGIN

  Query Match             1.1%;  Score 77.6;  DB 6;  Length 315;
  Best Local Similarity   56.2%;  Pred. No. 2.5e-09;
  Matches  146;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy       6313 CGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAG 6372
              ||   ||||    || |||  | ||| |     |  | | ||||  |||| |   |  |
Db        306 CGCACTTCCCATGCCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAA 247

Qy       6373 CACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCT 6432
               |  || ||| |  | || || ||||||||  | || |||||   ||| | |    ||
Db        246 GATGATGGCCGGGGCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTC 187

Qy       6433 GTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATG 6492
              |||||| ||||| ||     | ||||||  |||   |   ||||||||||  ||   |||
Db        186 GTGCGGCTCATGATGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATG 127
```

```
Qy       6493 CGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCA 6552
              |  |  |   || | ||  |||||| |||  | ||| || |  || |  || |   ||
Db        126 CATCGCCCAAGCGAAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGT 67

Qy       6553 TAACCAGAGCATTGTTCGTC 6572
                 ||||  |||    | ||
Db         66 CGGCCAGGACATGACCCATC 47


RESULT 4
CB042040/c
LOCUS       CB042040                 315 bp    mRNA    linear   EST 15-JAN-2003
DEFINITION  4007005 BARC-EMBRAPA 315BOV Bos indicus cDNA clone 315BOV_1P23
            Unknown, mRNA sequence.
ACCESSION   CB042040
VERSION     CB042040.1  GI:27761285
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 18
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: P  column: 23
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 315.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="315BOV_1P23"
                     /sex="female"
                     /tissue_type="mammary"
                     /cell_type="epithelium"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 315BOV"
                     /note="Organ: mammary; Vector: pUC118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature, PCR with 16mer AM5."
ORIGIN

  Query Match             1.1%;  Score 77.6;  DB 6;  Length 315;
  Best Local Similarity   56.2%;  Pred. No. 2.5e-09;
  Matches  146;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;
```

```
Qy       6313 CGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAG 6372
              ||   ||||    || |||  | ||| |     |  | | ||||  |||| |   |  |
Db        306 CGCACTTCCCATGCCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAA 247

Qy       6373 CACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCT 6432
               |  || ||| |  | || || ||||||||  | || |||||   ||| | |    ||
Db        246 GATGATGGCCGGGGCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTC 187

Qy       6433 GTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATG 6492
              |||||| ||||| ||     | ||||||  |||   |   ||||||||||  ||   |||
Db        186 GTGCGGCTCATGATGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATG 127

Qy       6493 CGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCA 6552
              |  |  |   || | ||  |||||| |||  | ||| || |  || |  || |   ||
Db        126 CATCGCCCAAGCGAAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGT 67

Qy       6553 TAACCAGAGCATTGTTCGTC 6572
                 ||||  |||    | ||
Db         66 CGGCCAGGACATGACCCATC 47


RESULT 5
CB041738/c
LOCUS       CB041738                 349 bp    mRNA    linear   EST 15-JAN-2003
DEFINITION  4006549 BARC-EMBRAPA 314BOV Bos indicus cDNA clone 314BOV_1D23
            Unknown, mRNA sequence.
ACCESSION   CB041738
VERSION     CB041738.1  GI:27760983
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 349)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: D  column: 23
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 349.
FEATURES             Location/Qualifiers
     source          1. .349
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="314BOV_1D23"
                     /sex="female"
                     /tissue_type="teat cistern"
                     /cell_type="epithelium"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 314BOV"
                     /note="Organ: mammary; Vector: pUC118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature, PCR with 16mer AM5 only."
ORIGIN

  Query Match             1.1%;  Score 77.6;  DB 6;  Length 349;
  Best Local Similarity   56.2%;  Pred. No. 2.6e-09;
  Matches  146;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy       6313 CGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAG 6372
              ||   ||||    || |||  | ||| |     |  | | ||||  |||| |   |  |
Db        340 CGCACTTCCCATGCCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAA 281

Qy       6373 CACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCT 6432
               |  || ||| |  | || || ||||||||  | || |||||   ||| | |    ||
Db        280 GATGATGGCCGGGGCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTC 221

Qy       6433 GTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATG 6492
              |||||| ||||| ||     | ||||||  |||   |   ||||||||||  ||   |||
Db        220 GTGCGGCTCATGATGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATG 161

Qy       6493 CGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCA 6552
              |  |  |   || | ||  |||||| |||  | ||| || |  || |  || |   ||
Db        160 CATCGCCCAAGCGAAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGT 101

Qy       6553 TAACCAGAGCATTGTTCGTC 6572
                 ||||  |||    | ||
Db        100 CGGCCAGGACATGACCCATC 81


RESULT 6
CB041616
LOCUS       CB041616                 359 bp    mRNA    linear   EST 15-JAN-2003
DEFINITION  4006367 BARC-EMBRAPA 314BOV Bos indicus cDNA clone 314BOV_1F09
            Unknown, mRNA sequence.
ACCESSION   CB041616
VERSION     CB041616.1  GI:27760861
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 359)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: F  column: 09
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 359.
FEATURES             Location/Qualifiers
     source          1. .359
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="314BOV_1F09"
                     /sex="female"
                     /tissue_type="teat cistern"
                     /cell_type="epithelium"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 314BOV"
                     /note="Organ: mammary; Vector: pUC118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature, PCR with 16mer AM5 only."
ORIGIN

  Query Match             1.1%;  Score 77.6;  DB 6;  Length 359;
  Best Local Similarity   56.2%;  Pred. No. 2.6e-09;
  Matches  146;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy       6313 CGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAG 6372
              ||   ||||    || |||  | ||| |     |  | | ||||  |||| |   |  |
Db         20 CGCACTTCCCATGCCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAA 79

Qy       6373 CACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCT 6432
               |  || ||| |  | || || ||||||||  | || |||||   ||| | |    ||
Db         80 GATGATGGCCGGGGCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTC 139

Qy       6433 GTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATG 6492
              |||||| ||||| ||     | ||||||  |||   |   ||||||||||  ||   |||
Db        140 GTGCGGCTCATGATGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATG 199

Qy       6493 CGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCA 6552
              |  |  |   || | ||  ||||| |||  | ||| || |  || |  || |   ||
Db        200 CATCGCCCAAGCGAAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGT 259

Qy       6553 TAACCAGAGCATTGTTCGTC 6572
                 ||||  |||    | ||
Db        260 CGGCCAGGACATGACCCATC 279


RESULT 7
CB061284
LOCUS       CB061284                 314 bp    mRNA    linear   EST 17-JAN-2003
DEFINITION  4011834 BARC-EMBRAPA 338BOV Bos indicus cDNA clone 338BOV_1H20
            Unknown, mRNA sequence.
ACCESSION   CB061284
VERSION     CB061284.1  GI:27799571
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 314)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: H  column: 20
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 314.
FEATURES             Location/Qualifiers
     source          1..314
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="338BOV_1H20"
                     /sex="female"
                     /tissue_type="teat cistern and alveoli"
                     /cell_type="multiple"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 338BOV"
                     /note="Organ: mammary; Vector: pUC 118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature. Cow AM1."
ORIGIN

  Query Match             1.1%;  Score 77.4;  DB 6;  Length 314;
  Best Local Similarity   57.1%;  Pred. No. 2.8e-09;
  Matches  141;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy       6326 CCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGT 6385
              || |||  ||||| |     |  | | ||||  |||| |   |  |  |   || ||| |
Db          4 CCGATCGGCTGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAAGATGATGGCCGGG 63

Qy       6386 ACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGG 6445
               | || || ||||||||  | || |||||   ||| | |    ||  |||||| |||||
Db         64 GCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTCGTGCGGCTCATGA 123

Qy       6446 TGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCT 6505
              ||     | ||||||  |||   |   ||||||||||  ||   ||||  |  |   ||
Db        124 TGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATGCATCGCCCATGCG 183

Qy       6506 ACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATT 6565
              | ||  |||||| |||  | ||| || |  || |  || |   ||     ||||  |||
Db        184 AAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGTCGGCCAGGACATG 243

Qy       6566 GTTCGTC 6572
                 | ||
Db        244 ACCCATC 250


RESULT 8
CB061218/c
LOCUS       CB061218                 300 bp    mRNA    linear   EST 17-JAN-2003
DEFINITION  4011666 BARC-EMBRAPA 338BOV Bos indicus cDNA clone 338BOV_1L08
            Unknown, mRNA sequence.
ACCESSION   CB061218
VERSION     CB061218.1  GI:27799505
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 300)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: L  column: 08
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 300.
FEATURES             Location/Qualifiers
     source          1..300
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="338BOV_1L08"
                     /sex="female"
                     /tissue_type="teat cistern and alveoli"
                     /cell_type="multiple"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 338BOV"
                     /note="Organ: mammary; Vector: pUC 118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature. Cow AM1."
ORIGIN

  Query Match             1.1%;  Score 75.8;  DB 6;  Length 300;
  Best Local Similarity   56.7%;  Pred. No. 7.7e-09;
  Matches  140;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       6326 CCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGT 6385
              || |||  | ||| |     |  | | ||||  |||| |   |  |  |  || ||| | 
Db        294 CCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAAGATGATGGCCGGG 235

Qy       6386 ACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGG 6445
               | || || |||||||||  | || ||||||   ||| | |    ||  |||||| ||||| 
Db        234 GCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTCGTGCGGCTCATGA 175

Qy       6446 TGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCT 6505
              ||     | ||||||  |||   |   ||||||||||  ||   ||||  |  |   || 
Db        174 TGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATGCATCGCCCAAGCG 115

Qy       6506 ACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATT 6565
              | ||  |||||| |||  | ||| || |  || |  || |   ||     ||||  ||| 
Db        114 AAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGTCGGCCAGGACATG 55

Qy       6566 GTTCGTC 6572
                 | ||
Db         54 ACCCATC 48


RESULT 9
CB059804/c
LOCUS       CB059804                 317 bp    mRNA    linear   EST 17-JAN-2003
DEFINITION  4009880 BARC-EMBRAPA 323BOV Bos indicus cDNA clone 323BOV_1B22
            Unknown, mRNA sequence.
ACCESSION   CB059804
VERSION     CB059804.1  GI:27798091
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 317)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: B  column: 22
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 317.
FEATURES             Location/Qualifiers
     source          1..317
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="323BOV_1B22"
                     /sex="female"
                     /tissue_type="parenchyma"
                     /cell_type="epithelium"
                     /dev_stage="involuted"
                     /lab_host="DH5alpha"
                     /clone_lib="BARC-EMBRAPA 323BOV"
                     /note="Organ: mammary; Vector: pUC 118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     sample from cow AM1, samples 2, 4, and 5 extracted on
                     3/27/02, RT with Superscript II at 37 deg C annealing
                     temperature, PCR with 16-mers."
ORIGIN

  Query Match             1.1%;  Score 75.8;  DB 6;  Length 317;
  Best Local Similarity   56.7%;  Pred. No. 7.8e-09;
  Matches  140;  Conservative    0;  Mismatches  107;  Indels    0;  Gaps    0;

Qy       6326 CCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGT 6385
              || |||  | ||| |     |  | | ||||  |||| |   |  |  |  || ||| | 
Db        311 CCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAAGATGATGGCCGGG 252

Qy       6386 ACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGG 6445
               | || || |||||||||  | || ||||||   ||| | |    ||  |||||| ||||| 
Db        251 GCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTCGTGCGGCTCATGA 192

Qy       6446 TGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCT 6505
              ||     | ||||||  |||   |   ||||||||||  ||   ||||  |  |   || 
Db        191 TGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATGCATCGCCCAAGCG 132
```

```
Qy       6506 ACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATT 6565
              | ||  |||||| |||  | ||| || |  || |  || |   ||     ||||  |||
Db        131 AAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCGTCGTCGGCCAGGACATG 72

Qy       6566 GTTCGTC 6572
                 | ||
Db         71 ACCCATC 65


RESULT 10
CB041601/c
LOCUS       CB041601                 354 bp    mRNA    linear   EST 15-JAN-2003
DEFINITION  4006361 BARC-EMBRAPA 314BOV Bos indicus cDNA clone 314BOV_1E09
            Unknown, mRNA sequence.
ACCESSION   CB041601
VERSION     CB041601.1  GI:27760846
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: E  column: 09
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 354.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="314BOV_1E09"
                     /sex="female"
                     /tissue_type="teat cistern"
                     /cell_type="epithelium"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 314BOV"
                     /note="Organ: mammary; Vector: pUC118; Site_1: HincII;
                     Site_2: HincII; This mammary-derived cDNA library was
                     created as part of a collaborative project between the ARS
                     Gene Evaluation and Mapping Laboratory and the EMBRAPA
                     Dairy Cattle Research Center under the sponsorship of
                     USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                     extracted on 6/8/02, RT with Superscript II at 37 deg C
                     annealing temperature, PCR with 16mer AM5 only."
ORIGIN

  Query Match            1.1%;  Score 74.4;  DB 6;  Length 354;
  Best Local Similarity  55.4%;  Pred. No. 1.9e-08;
  Matches  144;  Conservative    0;  Mismatches  116;  Indels    0;  Gaps    0;
```

```
Qy       6313 CGTCATTCCGGCACCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAG 6372
              ||   ||||    || |||  | ||| |     |  | | ||||  |||| |   |  |
Db        345 CGCACTTCCCATGCCGATCGGCAGCATCCAAGGCCAGACGTGCAACCCGAGAGCCACAAA 286

Qy       6373 CACAATCGCCAGTACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCT 6432
               |  || ||| |  | || || ||||||||  | || ||||    ||| | |    ||
Db        285 GATGATGGCCGGGGCGGCGAAAACCACCGCGAACAGGTCGTCCTTCTCCAGCACGTCGTC 226

Qy       6433 GTGCGGTTCATGGTGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATG 6492
              |||||| ||||| ||     | ||||||  |||   |   ||||||||||  ||   |||
Db        225 GTGCGGCTCATGATGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATG 166

Qy       6493 CGACAGCGCCGCTACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCA 6552
              |  |  |   || | ||  ||| || |||  | ||| || |  || |  || |   ||
Db        165 CATCGCCCAAGCGAAGACCTCCGTCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGT 106

Qy       6553 TAACCAGAGCATTGTTCGTC 6572
                 ||||  |||    | ||
Db        105 CGGCCAGGACATGACCCATC 86


RESULT 11
CB062064/c
LOCUS       CB062064                 302 bp    mRNA    linear   EST 17-JAN-2003
DEFINITION  4012128 BARC-EMBRAPA 339BOV Bos indicus cDNA clone 339BOV_1I14
            Unknown, mRNA sequence.
ACCESSION   CB062064
VERSION     CB062064.1  GI:27800351
KEYWORDS    EST.
SOURCE      Bos indicus (zebu)
  ORGANISM  Bos indicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
            Bovinae; Bos.
REFERENCE   1  (bases 1 to 302)
  AUTHORS   da Mota,A.F., Sonstegard,T.S., Van Tassell,C.P., Matukumalli,L.K.,
            Wood,D.L., Capuco,A.V., Brito,M.A.P., Martinez,M.L., Connor,E.E.,
            Machado,M.A. and Coutinho,L.L.
  TITLE     Construction and Characterization of cDNA Libraries Generated from
            Mammary Gland Tissues of Holstein (Bos taurus) and Gir (Bos
            indicus) Cattle
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Adilson F. da Mota
            Gene Evaluation and Mapping Laboratory
            USDA, ARS, Animal and Natural Resources Institute
            Bldg. 200 Rm3 BARC-East, Beltsville, MD 20705, USA
            Tel: 3015048456
            Fax: 3015048414
            Email: amota@cnpgl.embrapa.br
            Single pass sequencing. Bases called and trimmed with phred
            0.000925 using options -trim_alt '' -trim_fasta. Vector identified
            by cross_match using options -minmatch 12 -minscore 12
            PCR PRimers
            FORWARD: GTTTTCCCAGTCACGACGTTG
            BACKWARD: TGAGCGGATAACAATTTCACACAG
            Plate: 1  row: I  column: 14
            Seq primer: GTTTTCCCAGTCACGACGTTG
            High quality sequence stop: 302.
FEATURES             Location/Qualifiers
     source          1. .302
                     /organism="Bos indicus"
                     /mol_type="mRNA"
                     /strain="Brazilian Dairy Gir"
                     /db_xref="taxon:9915"
                     /clone="339BOV_1I14"
                     /sex="female"
                     /tissue_type="teat cistern and alveoli"
                     /cell_type="multiple"
                     /dev_stage="involuted"
                     /lab_host="K-12"
                     /clone_lib="BARC-EMBRAPA 339BOV"
                     /note="Organ: mammary; Vector: pUC 19; Site_1: SmaI;
                     Site_2: SmaI; This mammary-derived cDNA library was
```

```
                    created as part of a collaborative project between the ARS
                    Gene Evaluation and Mapping Laboratory and the EMBRAPA
                    Dairy Cattle Research Center under the sponsorship of
                    USDA, ARS/EMBRAPA-LABEX program in animal genomics. RNA
                    extracted on 6/8/02, RT with Superscript II at 37 deg C
                    annealing temperature. Cow AM1."
ORIGIN

  Query Match             1.1%;  Score 74.2;  DB 6;  Length 302;
  Best Local Similarity   56.3%;  Pred. No. 2.1e-08;
  Matches  139;  Conservative    0;  Mismatches  108;  Indels    0;  Gaps    0;

Qy       6326 CCAATCCACTGCAGCGGCCACATGCCTTGCACACCGACATAAATCAGCACAATCGCCAGT 6385
              || |||  | ||| |     |  |   ||||  |||| |   |  |  |  || ||| |
Db        296 CCGATCGGCAGCATCCAAGGCCAGATGTGCAACCCGAGAGCCACAAAGATGATGGCCGGG 237

Qy       6386 ACCGCAAACACCACCGCATAAAGATCGTTGAGCTCAAACTTACCGCTGTGCGGTTCATGG 6445
               | || || ||||||||  | || |||||   ||| | |    ||  |||||| |||||
Db        236 GCGGCGAAAACCACCGCGAACAGGTCGTTCTTCTCCAGCACGTCGTCGTGCGGCTCATGA 177

Qy       6446 TGCGACAGATGCCAGCCCCATCCCCAACCGTGCATGATGTATTTATGCGACAGCGCCGCT 6505
              ||     | ||||||  |||   |   ||||||||||  ||   ||||  |  |   ||
Db        176 TGGCTGCGGTGCCAGGTCCACAGCAGGCCGTGCATGACATAACGATGCATCGCCCAAGCG 117

Qy       6506 ACGATTTCCATCACCACCACGGTTGCCAACAAGATAAGCACGTTCCATAACCAGAGCATT 6565
              | ||  |||||| |||  | ||| || |  || |  || |   ||     ||||  |||
Db        116 AAGACCTCCATCCCCAGGAAGGTGGCGAGGAACAGCAGGATCATCGTCGGCCAGGACATG 57

Qy       6566 GTTCGTC 6572
                 | ||
Db         56 ACCCATC 50


RESULT 12
BH439236/c
LOCUS       BH439236                 475 bp    DNA     linear   GSS 12-DEC-2001
DEFINITION  BOHAC18TF BOHA Brassica oleracea genomic clone BOHAC18, genomic
            survey sequence.
ACCESSION   BH439236
VERSION     BH439236.1  GI:17624950
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 475)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOHAC18TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..475
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHAC18"
                     /clone_lib="BOHA"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             1.1%;  Score 73.6;  DB 8;  Length 475;
```

```
  Best Local Similarity   53.9%;  Pred. No. 3.5e-08;
  Matches  151;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        559 GCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATG 618
              || || || ||||||||| | ||| ||   |||||||| |  || ||  | || |  |||
Db        378 GCTTGCGCGGTGGAAATGATCCACACGATGTCGCTGATCCACGACGATCTCCCGTGTATG 319

Qy        619 GATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTG 678
              || ||||   |  | || || ||    || |||   ||    |  || || ||| ||||
Db        318 GACAACGACGACCTCCGCCGCGGAAAACCCACCAACCACAAAGTGTTCGGCGAAGACGTC 259

Qy        679 GCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCC 738
              ||| || | ||||    ||||||  |      ||| |||||     | ||     |  |
Db        258 GCGGTTTTAGCCGGAGACGCGCTTTTATCGTTCGCGTTTGAGCACTTGGCGTCGACGTCG 199

Qy        739 GGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTG 798
              ||  || || |            |  |||||| |  || ||  |    ||||| ||
Db        198 GGGGTGGCTCCGGCGAGGGTGGTTAGAGCGATCGGGGAGCTGGCGAGAGCCGTTGGATCA 139

Qy        799 CAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACG 838
               | ||  | |||   || ||  |   |||| |   |  ||
Db        138 GAAGGGCTTGTGGCGGGTCAGATTGTGGATATCAGCAGCG 99


RESULT 13
BZ434062
LOCUS       BZ434062                 754 bp    DNA     linear   GSS 13-DEC-2002
DEFINITION  BONFQ34TF BO_1.6_2_KB_tot Brassica oleracea genomic clone BONFQ34,
            genomic survey sequence.
ACCESSION   BZ434062
VERSION     BZ434062.1  GI:26683951
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 754)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BONFQ34TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..754
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BONFQ34"
                     /clone_lib="BO_1.6_2_KB_tot"
                     /note="Vector: pHOS1; Site_1: BstXI; 1.6-2 kb sheared
                     total DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             1.1%;  Score 73.6;  DB 8;  Length 754;
  Best Local Similarity   53.9%;  Pred. No. 3.9e-08;
  Matches  151;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy        559 GCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATG 618
              || || || ||||||||| | ||| ||   |||||||| |  || ||  | || |  |||
Db        194 GCTTGCGCGGTGGAAATGATCCACACGATGTCGCTGATCCACGACGATCTCCCGTGTATG 253

Qy        619 GATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTG 678
              || ||||   |  | || || ||    || |||   ||    |  || || ||| ||||
Db          254 GACAACGACGACCTCCGCCGCGGAAAACCCACCAACCACAAAGTGTTCGGCGAAGACGTC 313

Qy          679 GCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCC 738
                ||| || | ||||    ||||||  |      ||| |||||     | ||     |  |
Db          314 GCGGTTTTAGCCGGAGACGCGCTTTTATCGTTCGCGTTTGAGCACTTGGCGTCGACGTCG 373

Qy          739 GGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTG 798
                ||  || || |            |  |||||| |  || ||  |    ||||| ||
Db          374 GGGGTGGCTCCGGCGAGGGTGGTTAGAGCGATCGGGGAGCTGGCGAGAGCCGTTGGATCA 433

Qy          799 CAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGACG 838
                 | ||  | |||   || ||  |   |||| |   |  ||
Db          434 GAAGGGCTTGTGGCGGGTCAGATTGTGGATATCAGCAGCG 473


RESULT 14
BE323612
LOCUS       BE323612                 425 bp    mRNA    linear   EST 21-DEC-2000
DEFINITION  NF006A09PL1F1065 Phosphate starved leaf Medicago truncatula cDNA
            clone NF006A09PL 5', mRNA sequence.
ACCESSION   BE323612
VERSION     BE323612.2  GI:11967172
KEYWORDS    EST.
SOURCE      Medicago truncatula (barrel medic)
  ORGANISM  Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE   1  (bases 1 to 425)
  AUTHORS   Liu,J., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
            Flores,H.R., Inman,J.T., Weller,J.W., May,G.D. and Harrison,M.J.
  TITLE     Expressed Sequence Tags from the Samuel Roberts Noble Foundation
            Medicago truncatula phosphate-starved leaf library
  JOURNAL   Unpublished (2000)
COMMENT     On Jul 14, 2000 this sequence version replaced gi:9197389.
            Contact: Harrison MJ
            Plant Biology Division
            The Samuel Roberts Noble Foundation
            2510 Sam Noble Parkway, Ardmore, OK 73402, USA
            Tel: 580 221 7325
            Fax: 580 221 7380
            Email: mjharrison@noble.org
            Medicago Genome Initiative accession: MGI:S:20177
            Insert Length: 818   Std Error: 0.00
            Plate: 006  row: A  column: 09
            Seq primer: TCACACAGGAAACAGCTATGAC.
FEATURES             Location/Qualifiers
     source          1..425
                     /organism="Medicago truncatula"
                     /mol_type="mRNA"
                     /db_xref="taxon:3880"
                     /clone="NF006A09PL"
                     /tissue_type="leaf"
                     /dev_stage="trifoliate"
                     /clone_lib="Phosphate starved leaf"
                     /note="Vector: Lambda Zap; At the trifoliate stage, M.
                     truncatula plants were transplanted to phosphate-free sand
                     and grown for a further 30 days.  During this 30 day
                     period, the plants were fertilized twice weekly with 1/2
                     Hoaglands solution containing only 20uM potassium
                     phosphate.  RNA was prepared from above ground tissues."
ORIGIN

  Query Match             1.0%;  Score 72;  DB 2;  Length 425;
  Best Local Similarity   62.0%;  Pred. No. 9.2e-08;
  Matches  114;  Conservative    0;  Mismatches   70;  Indels    0;  Gaps    0;

Qy          558 CGCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCGAT 617
                |||||| || || |||||| | ||| ||   || || || |  |||||| |||| |  ||
Db          233 CGCCTGCGCCGTTGAAATGATCCACACGATGTCTCTCATCCACGATGACCTTCCTTGTAT 292

Qy          618 GGATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGT 677
```

```
                ||||||||   |  | || || |||   |||||    ||    |  || || ||| ||||
Db          293 GGATAACGATGATCTCCGGCGAGGTAAACCTACAAACCACAAAGTCTTCGGAGAAGACGT 352

Qy          678 GGCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACC 737
                 ||  |||||||||     || || ||       || || |||   |||||| |  |  |
Db          353 TGCTGTTCTCGCCGGAGATGCTCTTCTCGCTTTTGCTTTCGAACATATTGCCGTCTCCAC 412

Qy          738 CGGT 741
                || |
Db          413 CGTT 416


RESULT 15
BH502634
LOCUS       BH502634                 709 bp    DNA     linear   GSS 13-DEC-2001
DEFINITION  BOHJK46TR BOHJ Brassica oleracea genomic clone BOHJK46, genomic
            survey sequence.
ACCESSION   BH502634
VERSION     BH502634.1  GI:17710731
KEYWORDS    GSS.
SOURCE      Brassica oleracea
  ORGANISM  Brassica oleracea
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE   1  (bases 1 to 709)
  AUTHORS   Town,C.D., Van Aken,S., Utterback,T., Koo,H. and Fraser,C.M.
  TITLE     Whole genome shotgun sequencing of Brassica oleracea
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: BOHJK46TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            DNA is from a doubled haploid provided by Tom Osborn.
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1..709
                     /organism="Brassica oleracea"
                     /mol_type="genomic DNA"
                     /strain="TO1000DH3"
                     /db_xref="taxon:3712"
                     /clone="BOHJK46"
                     /clone_lib="BOHJ"
                     /note="Vector: pHOS1; Site_1: BstXI; 2-3 kb sheared
                     genomic DNA inserted into pHOS1 using BstXI linkers"
ORIGIN

  Query Match             1.0%;  Score 72;  DB 8;  Length 709;
  Best Local Similarity   50.2%;  Pred. No. 1.1e-07;
  Matches  239;  Conservative    0;  Mismatches  225;  Indels   12;  Gaps    2;

Qy          559 GCCTGTGCAGTGGAAATGGTGCACGCGGCATCGCTGATTCTGGATGACATTCCCTCGATG 618
                || || || || || ||| | ||| ||   || || ||    || ||  |||| |  |||
Db           87 GCTTGCGCGGTCGAGATGATTCACACGATGTCACTCATCAAAGACGATCTTCCTTGCATG 146

Qy          619 GATAACGCGCAGATGCGTCGTGGTCGCCCTACCGTGCATCGCGAATTTGGTGAAAACGTG 678
                || ||||   |  | || || |||   || ||   |||    |  || || |||| |||
Db          147 GACAACGACGACCTCCGGCGAGGTAAGCCCACGACGCACAAAGTCTTCGGAGAAAGCGTC 206

Qy          679 GCGATTCTCGCCGCCATCGCGCTGCTTAGCCGCGCATTTGAAGTGATTGCCATTGCACCC 738
                ||||||||| ||| |  |||||| ||    |  || || ||     |  |    ||   |
Db          207 GCGATTCTCTCCGGCGGCGCGCTCCTAGCTCTGGCCTTCGAGCATTTGACGGAAGCTGAC 266

Qy          739 GGTTTGCCTGCCATACATAAATCTGAAGCGATTGCTGAACTCTCCGCTGCCGTCGGCCTG 798
                |  | |||||  | |   |    |  |||| ||   |||||  |     |  | ||   |
Db          267 GTGTCGCCTGAGAGA---ATGGTTAGAGCGGTTAAAGAACTGGCGAAGTCTATAGGGACG 323

Qy          799 CAGGGCTTAGTGCAAGGGCAATTCCAGGATCTGCACGAC---------GGCACGCAGAGC 849
                 | ||  | |||   || |||     |||| ||  |  |         |      |   |
Db          324 AAAGGGCTCGTGGCGGGACAAGCGATGGATTTGAGCAGCGAAGGTTTGGATCAAAACGAC 383

Qy          850 CGCAGCCCGGAAGCGATCGCCATGACCAACGAACTGAAAACCAGCGTGCTGTTTCGCGCC 909
                  | |   | | | | | |   | |   |||  |  |||||| |   |||| ||   ||
Db          384 GTCGGGTTGAAGGAGCTTGAGTTTATTCACGTTCATAAAACCGGTTCGCTGCTTGAGGCT 443

Qy          910 ACGCTGCAAATGGCGGCGATTGCCGCTGACGCTTCACCGCAGGTGCGGCAAAGACTTAGC 969
                 ||  |   || |  ||| ||   |  |  | |||   | | | |    | |   | ||
Db          444 TCGGCGGTTATCGGAGCGGTTATTGGAGGTGGTTCGGAGGAAGAGGTTGAGAAGGTGAGG 503

Qy          970 TTCTTCGCCCAGGATTTGGGCCAGGCGTTTCAACTGCTCGACGACCTCGCCGACGG 1025
                   |||||   |  | | ||   |  ||||||  || | || ||  |    || ||
Db          504 AGGTTCGCGAGGTGTATTGGGTTGTTGTTTCAGGTGGTTGATGATATTTTGGATGG 559

Search completed: January 25, 2005, 10:29:04
Job time : 19364 secs
```